DEMOCRACY REPORTING INTERNATIONAL

DECEMBER 2021 • APRIL 2022

ONLINE PUBLIC DISCOURSE
IN THE MENA REGION

2022 LEBANESE AND JORDANIAN ELECTIONS AS A CASE STUDY

This first report has been produced by DRI and its partners for the project "Words Matter". The report covers the period from December 2021 to April 2022:

DRI Partners

Supported by

September 2022

Acknowledgments

Institut de Presse et des
Sciences de l'Information (IPSI)
– Tunisia

Association réseau
Mourakiboun
– Tunisia

Sudanese
Development
Initiative
(SUDIA)
– Sudan

DRI would like to earnestly thank
its regional partners for their
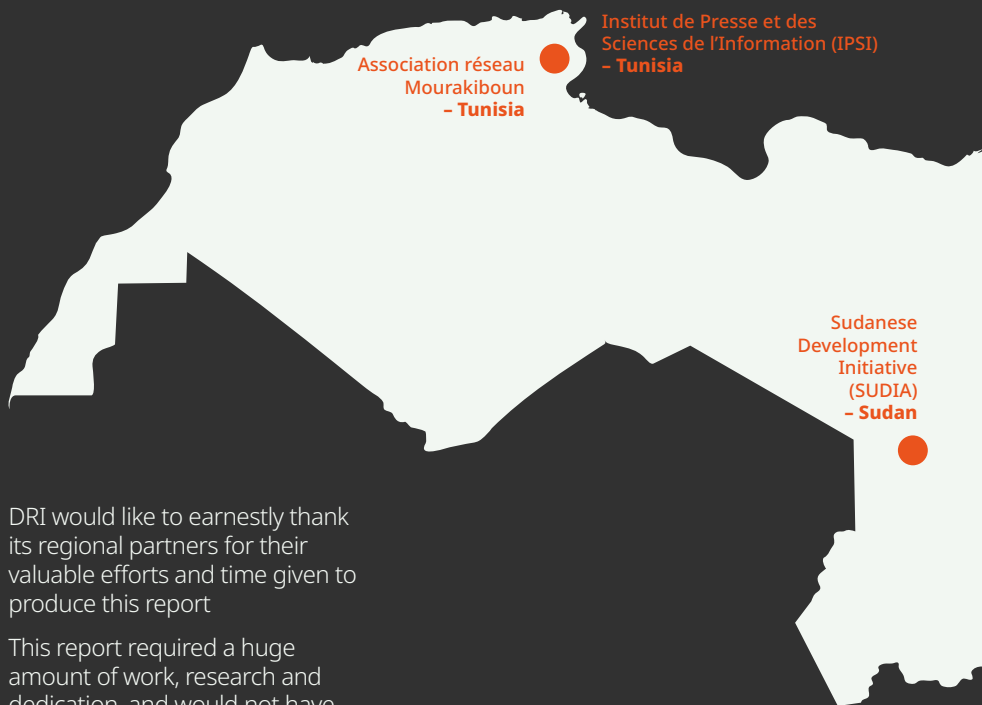valuable efforts and time given to
produce this report

This report required a huge
amount of work, research and
dedication, and would not have
been possible if we did not have
the support of many individuals
and organisations.

We would like, therefore, to
extend our sincere gratitude to
all of them.

Index

NB: Please check the Arabic version of the report to get introduced to the Arabic version of the social media monitoring toolkit.

Maharat Foundation **– Lebanon**

Al Hayat Center - RASED **– Jordan**

Abstract

Within the framework of the "Words Matter" project, Democracy Reporting International (DRI) and its partners are seeking to contribute to the strengthening of safeguards for democratic processes and of society's resilience to online disinformation and hate speech in the MENA region.

This report is the result of the contributions of DRI's partners in Tunisia, Jordan, Lebanon and Sudan, who provided data-driven content related to their specific contexts from December 2021 to March 2022. It is the first report to be published as part of "Words Matter", and will be followed by three other reports.

The report aims to:

- Analyse disinformation and hate speech during key national democratic processes, in order to shed the light on the behaviours, patterns and streams that are observed during these processes; and

- Develop national and regional recommendations for transparent regulations to combat online disinformation and hate speech.

With different levels of their progress in each country, the report presents the work accomplished by DRI's partners as follows:

First, it presents a case study of hate speech on Facebook during the municipal elections in Jordan, focusing on the election campaign silence period starting on 21 March 2022, and on election day, 22 March 2022. Our partner Al Hayat Centre carried out an analysis of posts and comments identified on Facebook local media platforms to provide evidence of the types and intensity of hate speech. A set of recommendations dedicated to national authorities and civil society actors in Jordan were formulated accordingly.

Second, it presents results from a report of monitoring during Lebanon's 15 May 2022 parliamentary elections, focusing on the political landscape before election day, from February 2022 to the end of March 2022. This monitoring, conducted by the Maharat Foundation, features the topics most discussed either by political candidates or influencers. It also contains a case study of the online hate speech campaign against the Lebanese journalist Dalia Ahmed, and provides national recommendations to political actors and the government. The next report will cover the period between April 2022 to election day, and then after election day, to provide the full picture of this important political process.

Third, the "LAB'TRACK" project presents its context, focus and methodology. This project joins the efforts of Mourakiboun and the Institute of Press and Information Sciences, who aim to study and understand the behaviour behind the online political disinformation phenomenon in Tunisia in the aftermath of the 25 July 2022 referendum.

Lastly, the Sudanese Development Initiative (SUDIA) studies the hate speech and disinformation phenomena in Sudan in the aftermath of the military coup of 25 October 2022. In this report, SUDIA examines the Sudanese political and social contexts, and outlines their methodology to address these phenomena.

Executive Summary

This report is the first of four regional social media monitoring reports that will be produced under the DRI "Words Matter" project, focusing on countering disinformation and hate speech in the MENA region. The project aims to strengthen the safeguarding of democratic processes and society's resilience to online disinformation and hate speech in the MENA region. The project builds on the assumption that civil society actors, including journalists and media, are essential to monitoring, understanding, and raising awareness of what debates and discourses are occurring online. The project consists of three main components:

- Capacity-building for civil society organisations (CSOs) allowing them to acquire institutional skills to design sound social media monitoring methodologies, to effectively monitor disinformation and hate speech online, and to enhance evidence of the impacts of disinformation and hate speech online on political participation and human rights.

- Enhanced multi-stakeholder and regional engagement, to advocate against and combat online disinformation and hate speech through a civil society network, as well as continuous exchange on transparent regulations; and

- An improved awareness and resilience of civic target groups and concrete action by decision-makers to transparently combat hate speech and disinformation online.

The project operates in four MENA countries: Tunisia, Sudan, Lebanon, and Jordan. Our partners in the regional project include L'Institut de Presse et des Sciences de l'Information (IPSI) and Mourakiboun, from Tunisia, the Sudanese Development Initiative (SUDIA), from Sudan, the Maharat Foundation, from Lebanon, and the Al-Hayat Center for Civil Society Development and Jordan Open-Source Association (JOSA), from Jordan.

The report investigates online disinformation and hate speech trends during key national democratic processes, to shed the light on the

behaviours, patterns and streams that are observed during these processes. It also develops national and regional recommendations aimed towards CSOs, researchers and social media platforms, to promote transparent regulations to combat online disinformation and hate speech. The report presents the results of social media monitoring efforts and the methodologies used during elections in two countries (Jordan and Lebanon) and showcases the methodologies that will be used by our partners in the future in Tunisia and Sudan.

To address the lack of data analysis and training resources in Arabic, DRI localised its English version of the Digital Democracy Monitor Toolkit and produced an Arabic version of the Toolkit, to make it more relevant to the MENA region context. The Toolkit is structured into three phases – preparation, data analysis and reporting – to help journalists and researchers conduct social media monitoring on their own. It is one of the first practical social media monitoring toolkits available in Arabic to help civil society, journalists and academia research social media and online democracy without outsourcing the technical aspects of the work to international organizations.

The Arabic localization of the Toolkit has proven to be a valuable resource to our partners in building a strong

methodology and research approach in their projects. For example, Maharat, our Lebanese partner, used it to monitor hate speech in parliamentary election campaigns, and have also used other resources and classifications in other DRI guides in their work on gender-based violence. SUDIA, our Sudanese partner, adapted the toolkit to guide their methodology to monitor hate speech and disinformation during the transition phase after 25 October military coup in Sudan.

The report includes findings from two major events that were covered in the social media monitoring efforts up until March 2022. These are the municipal elections in Jordan, which took place in March, and the online campaigning preliminary to the parliamentary elections in Lebanon, which were held in May.

In Lebanon, the Maharat Foundation analysed posts from a sample of political candidates running for the 2022 national elections and found that most messages played on the emotions of citizens (257 of 522 tweets/posts observed) or were accusations (173 of 522 tweets/posts). Conspiracy theory posts (34 of 522 tweets/posts) came a distant third. According to Maharat, 94.2 per cent of tweets and posts could be described as "populist", aimed at dissuading voters from calling for reforms and responding to societal needs. This

applies equally to content coming candidates from emerging political movements, who are supposed to support reform projects and offer rational criticism of the forces they seek to replace. Influencers and partisans of political groups took the same approach, by adopting their leaders' discourses and defending their positions, while also spreading rumours and disinformation. The analysts also found instances of accounts sharing manipulated news, and the report features the case of Dalia Ahmed, a Lebanese television host, who was the victim of an online harassment campaign. The report also demonstrates that political campaigning in Lebanon is managed by anonymous page administrators, which could raise concerns about transparency. The same concern applies to pages administered outside the territory of the country in question.

In Jordan, after analysing 51 Facebook media pages during the 2022 local elections, the Al Hayat Center found that 23.4 per cent of the 11,255 comments analysed contained hate speech. Out of these, defamation was the highest form of hate speech, accounting for 25.78 per cent of all hateful comments, followed by denigration (25.4 per cent), cyberbullying (18.97 per cent), and insults (10.72 per cent).

This report includes country-specific and regional recommendations, based on the social media monitoring efforts and observations from our partners in the four countries. At the regional level, in order to enhance information integrity, so citizens can form their own opinions and vote without being exposed to manipulation campaigns, as well as to prevent hate speech from being translated into violence, the report recommends:

- Agreeing on a definition of hate speech that does not curb the freedom of expression online; one of the main challenges for our partners during their social media monitoring proved to be agreeing on one definition of hate speech that takes into consideration their local contexts, while respecting international standards.

- Building regional networks and coalitions to facilitate the exchange of knowledge, including about the main behaviours related to disinformation and hate speech in the countries of the partner organizations.

- Gathering data and encouraging open data practices, while respecting personal data protection standards. This is to be done by working with social media platforms so that they provide greater access to social media monitoring tools

to CSOs and researchers in the MENA region, and work with the platforms on identifying hate speech and disinformation trends in the region, while ensuring high standards from regional actors to protect personal data; and

• Building better coordination mechanisms with social media platforms on content removal, where, according to our partners' experience, the detection of harmful content becomes very challenging when social media platforms remove content without providing a clear explanation for the reasons of removal.

For country-specific recommendations, the analysis undertaken by DR and its partners also examines the legislative framework of each country. All of these use imprecise definitions of freedom of speech online in their legislations, which could lead to overreaching restriction on this freedom. They also lack consistency in applying rules and regulations governing online speech. Aside from amending legislation to address hate speech and better protect freedom of expression, the project recommends the following:

For Jordan

• Clearly define and regulate political campaigning on social media.

• Introduce explicit penalties and sanctions in current legislation for those that commit bullying and electronic violence.

• Provide shorter and simplified legal procedures against those who use hate speech.

• Collaborate with social media platforms available in Jordan to develop policies that tackle hate speech and false and misleading content.

• Include information on hate speech and disinformation monitoring mechanisms in the curriculum of Media and Journalism Colleges, as well as build an anti-hate speech strategy, by educating their students about hate speech and methods to counter it, and how to contribute to raising public awareness of all forms of hate speech; and

• Promote programmes and initiatives by the government and CSOs to counter disinformation and hate speech

For Lebanon

• Enhance collaboration between CSOs working on social media monitoring and social media platforms to combat disinformation and hate speech more effectively.

• Increase digital media literacy among different social media

users to detect and combat online hate speech, and to report it using the proper mechanisms.

- Build the fact-checking capacities of journalists and CSOs; and

- Create a culture of accountability for political candidates

The report also contains a case study on gender-based violence on Twitter in the MENA region. The study was conducted by Helmi Noman, a social media researcher and analyst, who has examined the digital landscape in the MENA region at the Berkman Klein Center for Internet & Society, Harvard University. The study investigates to what extent politically active women in the MENA region face online gender-based violence on the Twitter platform. The study includes four highly influential and politically active women on Twitter: Loujain Hathloul, a Saudi women's rights activist and a political prisoner; Dima Sadek, a Lebanese television news anchor; Ghada Oueiss, a principal news presenter for Al Jazeera TV from Lebanon; and Tawakkol Karman, a Yemeni Nobel Peace Prize Laureate. The study finds that all four women have been targeted by offensive content (text and imagery) due to their public comments on contentious political issues. The offensive content was generated by a few accounts and was amplified by retweets and replies. Although a counternarrative against online GBV

has emerged within this discourse, the offenders used counter hashtags to disseminate more offensive tweets and amplify their reach.

Broader, contextualised research is still needed to explore hate speech, in general, and GBV in the Arabic Twitter sphere and on other social media platforms, and to understand to what extent this phenomenon causes women in the region to self-censor themselves in the public space, or to withdraw completely from these platforms. CSOs should also continue to engage Twitter in reviewing and acting on GBV, and to hold them responsible for their inaction. For its part, Twitter should invest more in detecting hate content in Arabic, and in its various dialects and contexts, and address its West-centric bias.
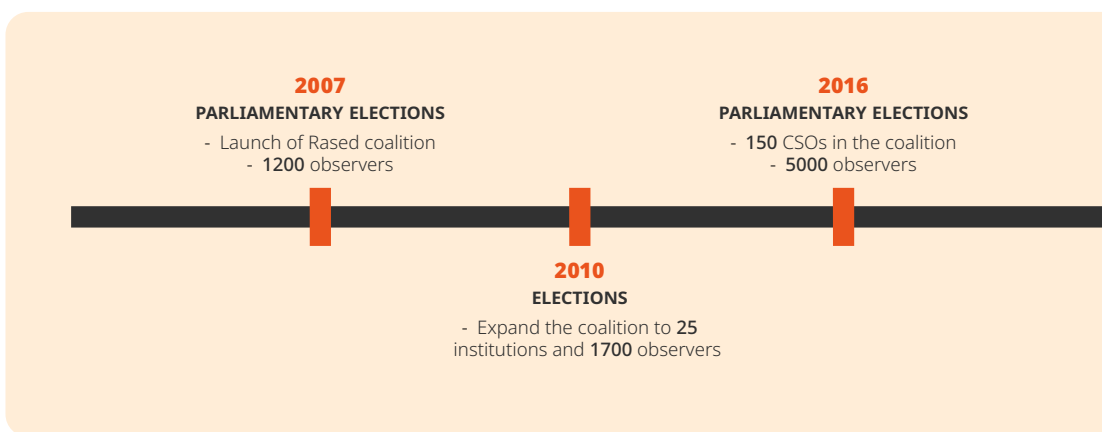
Country-based Analysis

1. Hate speech in Facebook on election silence: A case study from Jordan

1.1. About the Al Hayat Centre - RASED

The Al Hayat Centre-RASED is a CSO. It aims to promote accountability, governance, public participation and tolerance in Jordan and the region, within the framework of democracy, human rights, active citizenship, and the rule of law, while taking into consideration gender mainstreaming in public policy and actions. Al-Hayat works through five main programmes: (1) electoral reform RASED "the observer"; (2) governmental reform; (3) parliamentary reform; (4) local administration reform; and (5) social cohesion.

RASED's coalition for election observation and monitoring was established in 2007, during the parliamentary elections that year. The coalition was launched by 1,200 observers, who were deployed outside

2007
PARLIAMENTARY ELECTIONS
- Launch of Rased coalition
- **1200** observers

2010
ELECTIONS
- Expand the coalition to **25** institutions and **1700** observers

2016
PARLIAMENTARY ELECTIONS
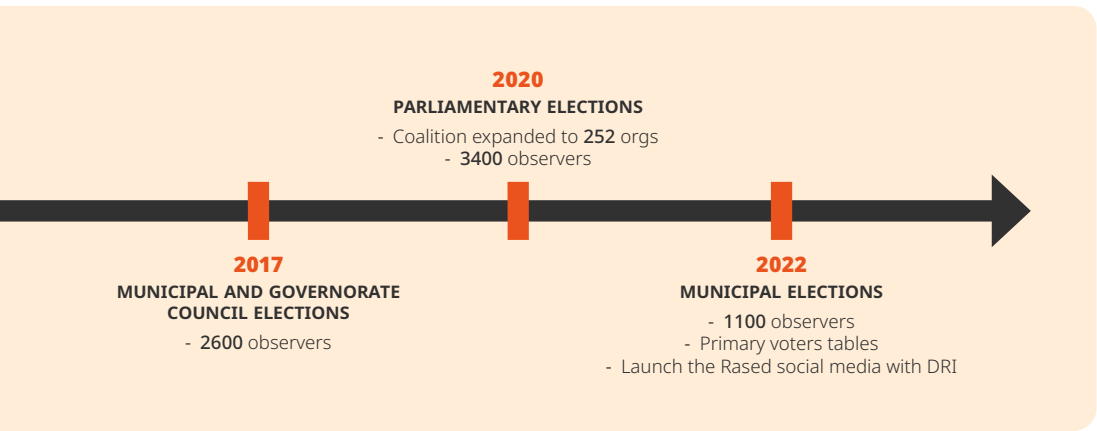- **150** CSOs in the coalition
- **5000** observers

of voting centres to observe the progress of the electoral process for the first time in Jordanian election history.

RASED's team continued working to enshrine its approach of election observation and codification within the Jordanian democratic system in 2010, when the team was allowed to bring 1,700 observers to voting centres to observe and monitor the elections.

In 2017, 2,600 observers took part in the RASED coalition's observation of the 2017 municipal and governorate council elections. In 2020, the Coalition's work was expanded to 252 organisations spread across all constituencies. Moreover, RASED's coalition observed the parliamentary elections, with 3,400 women and men observers taking part within the coalition. In addition, the municipal and governorate councils' elections were observed in 2022, with 1,100 women and men observing all of the electoral processes, from the publication of primary voters' tables until the results were announced. The coalition is not a governmental entity or political party, as it operates independently and impartially.

2020
PARLIAMENTARY ELECTIONS
- Coalition expanded to **252** orgs
- **3400** observers

2017
MUNICIPAL AND GOVERNORATE COUNCIL ELECTIONS
- **2600** observers

2022
MUNICIPAL ELECTIONS
- **1100** observers
- Primary voters tables
- Launch the Rased social media with DRI

1.2. Context

Jordan passed its first decentralisation law in 2015 and held its first decentralisation elections in 2017. The elections are for positions on governorate councils (newly created) and municipal councils, as well as mayoralties and the council for the Greater Amman municipality. The elections highlight Jordan's efforts to give more voice to citizens and increase their involvement in developmental decision-making processes. Based on this, the importance of monitoring activity around these elections on social media lies in the fact that it deals with sub-national politics. This sheds light on hate speech and disinformation trends at the governorate and municipal levels and provides insights into certain geographical areas and how trends there differ from those at the national level.

The final list of candidates for municipal elections and for the Greater Amman municipality was published on February 2022, and included 3,953 men and 867 women, according to statistics shared by IEC. As for the results, they were announced by the IEC and published in the Official Journal on 24 March 2022.

Within the framework of studying and monitoring hate speech in Jordan, the Al-Hayat Centre - RASED, with the support of DRI, published its report.

This is one of several reports that the Centre is working on through the «Monitoring Hate Speech on Social Media Platforms - RASED» project. The report monitors content related to municipal elections and governorates councils on social media platforms, following a particular methodology for monitoring hate speech and false news on those on which the Centre worked.

It focused on monitoring the Facebook platform on 21 and 22 March 2022, as well as 51 web pages that were selected according to specific criteria and to keywords such as: "elections, governorate councils, Independent Electoral Commission (IEC), RASED, polling percentage, ballot boxes, number of voters, election results, tabulation and counting results», and on posts that received four comments or more.

The monitoring during this period identified a growing number of phenomena and behaviours, including defamation.

1.3. Methodology

The research team conducting the monitoring and analysis were Al-Hayat Centre employees.

The methodology consists of three parts:

1.3.1. Purpose of the Study and Research Design

The observation of hate speech and disinformation on social networking sites was based on a clear research and analysis methodology, including a set of qualitative and quantitative indicators, developed based on good practices in this area and on the DRI

Toolkit. The methodology contained a set of local indicators to define hate speech and its classifications, and to indicate the extent of the distinction between criticism that should be protected by freedom of expression and hate speech. including in legislation that contains these two terms.

There is no definition or clarity in international law of what is hate speech. Jordan's legislation also lacks a comprehensive and clear definition of hate speech. When reviewing relevant Jordanian legislation, the Social Media Monitoring (SMM) team adopted the United Nations definition as a reference for this report.

The definition identifies **hate speech** [2] as «[a]ny type of verbal, written or behavioural communication which attacks or uses a pejorative or discriminatory language by reference to a person or group of people on the basis of identity, religion, nationality, race, colour or origin».

While speaking of **freedom of expression** [3] as a term, and based on the provisions of the Jordanian Constitution, specifically, article (15), which "[o]bliges the Country to ensure the freedom of expression and opinion of Jordanians", this freedom is absolute and unrestricted in form, time and means, However, the Constitution stipulates that the limits of the law shall not be exceeded. According to the laws, the most important

determination is that the right to express one's opinion does not extend to attacks on the honour, reputation or beliefs of others, or on the security and stability of the State.

Also, article 19 of the International Covenant on Civil and Political Rights, concerning the Right to Freedom of Expression and Opinion, provides the following consensual definition, which will be relied upon in the next methodology for action and research:

> *Everyone has the right to express his or her views, ideas, and beliefs freely, and everyone has the right to adopt and express views without prejudice, regardless of how it is used, whether it is in direct contact with people or by writing, radios, newspapers and other media, with the need to respect or hear the rights and beliefs of others, Without prejudice to national security, public order, public health or morals.*

---

[2] United Nations, "The United Nations Strategy and Action Plan Regarding Hate Speech", May 2019.

[3] "The Constitution of the Hashemite Kingdom of Jordan", p. 3.

1.3.2. Methods

After agreeing on a clear definition of hate speech and freedom of expression, Al Hayat began designing the research process, by identifying hate speech classifications and their intensity, in addition to establishing a clear data analysis procedure, from identifying the date sources through to the final findings.

Hate Speech Classifications

The SMM team worked on sorting hate speech into ten main classifications, based on international best practices and in line with the Jordanian context, laws and regulations. The definitions of the classifications are provided below:

TERMS	DEFINITIONS
Violence	Any violent action motivated by an individual's aggressiveness and likely to result in physical, sexual, psychological, digital or other harm or suffering, including threats of such acts, or coercion or arbitrary deprivation of liberty, whether occurring in public or private life.
Bullying	Bullying is a form of aggressive behaviour that is deliberately and repeatedly carried out by a person to insult or annoy another person. Bullying can take various forms, including physical bullying, verbal bullying or bullying with aggressive acts. This is undesirable behaviour, involving psychological dysfunction and disparities in power and social standing, and those bullied often suffer from subsequent psychological problems.
Insults	An attack on the dignity, honour, or dignity of others, verbally or in writing.
Incitement	The creation of a determination in others to commit and carry out any crime or act of violence. In other words, it creates the idea of committing acts of violence or crime in in another or others, with the intention of inducing them to commit such acts.
Sexual harassment	Any unwelcome sexual act, such as physical contact, sexual comments, the display of pornography or sexual requests, whether by word, deed or gesture. This act may be humiliating and may lead to physical and/or psychological harm.
Denigration	Any public statement, writing, drawing, picture or any sign or expression that diminishes the respect enjoyed by the party to which it is addressed.

Discrimination	Any distinction, exclusion, restriction or preference based on race, colour, descent, or national or ethnic origin that has the purpose or effect of nullifying or impairing the recognition, enjoyment or practice, on an equal footing, of fundamental freedoms in the political, economic, social or cultural fields, or in any other field of public life.
Defamation	The false attribution of an event or an incident in violation of the law and the well-established traditions in the country to a person, requiring punishment and general social contempt for that person. This attribution is public and deliberate and harms the reputation of the person or institution that it targets. This may be defamation by some people publishing personal pictures in an immoral way on social networking sites or publishing false news about them and working to publish them broadly, with the intention of disseminating the news to as many people as possible.
Disinformation	Information intended to cause harm to another individual or group that is knowingly false. This information takes several forms: – False stories posted on websites – Fabricated photographs and video clips – Information, photographs, and quotations being presented outside of their actual context (e.g., time or place).
Rumours	False news that spreads in the community quickly and is believed to be true. This false news is always interesting and arouses the curiosity of society and researchers, and these rumours usually lack a trusted source or verification of their authenticity.

Table 1: Hate Speech Classifications
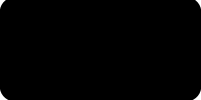
Figure 1: Screenshot of a rumour published in a Facebook post during the monitoring by Al Hayat

Hate Speech Intensity Scale

While hate speech represents a single category, it is important to note that there are clear differences in the intensity of such speech, and these vary from one context to another and from one case to another.

The SMM team, after reviewing the literature and best practices, established the following hate speech intensity Scale:

RATE [4]	INTENSITY CLASSIFICATION	DESCRIPTION	COLOUR-CODING
0	No Intensity	Content that is not related to the post and does not incite in any way.	
1	Disagreements	An expression that reflects a difference of opinion in relation to an idea, belief, etc.	
2	Negative Actions	An expression containing non-violent actions associated with a group or party, or responses containing non-violent actions, such as metaphors. Examples include accusations of theft, threats, indecency, mistreatment and alienation.	
3	Building a negative character	An expression containing a non-violent characterisation and insults, such as accusations of stupidity, robbery, counterfeiting, insanity,.	
4	Demonizing and Dehumanising	An expression containing inhumane and characterisation of inferiority, such as the use of labelling words associated with animals, diseases and others.	
5	Violence	An expression that involves inflicting physical or metaphorical harm, inciting such harm, and responses that call for physical or metaphorical violence, such as torture, rape, beatings, etc.	
6	Death	An expression that includes the word "murder" by a particular group, and responses that involve murder.	

Table 2: Hate speech Intensity scale [5]

---

[4] 1 is the lowest and 6 is the highest.

[5] Babak Bahador, "Classifying and Identifying the Intensity of Hate Speech", Items, 17 November 2020.

1.3.3. Statistical Data Analysis Procedure

The statistical data analysis procedure was divided into the following steps:

First, Identifying the Scope of Work:

The SMM team identified the scope of the work to focus on hate speech on the social media platform Facebook during the local administration elections in Jordan in March 2022, with a time frame of 21 and 22 March.

Second, Identify the Data Sources:

The SMM team prepared a list of 51 (governmental, and non-governmental) Facebook pages, chosen based on one or more of the following criteria:

1. The page should be Facebook verified and have a broad reach.

2. The owner of the page is a public figure.

3. The page belongs to the Jordanian government, ministries, and independent bodies, or is the official page of a Jordanian public or private university.

4. The page belongs to visual, audio, or written media that have a high number of followers and broad reach.

Third, Choosing the data analysis tool:

The SMM team, with technical help provided by the DRI team, chose the CrowdTangle tool to analyse the data. This method collects the data based on its sources, as well as on a variety of keywords.

Fourth, Identifying the Keywords:

The keywords chosen to collect the data were:

Elections, the elections, Governorate councils, independent electoral commission, RASED, voting percentage, voting boxes, numbers of voters, election results, sorting and sorting results.

The keywords were used in the Arabic language [6].
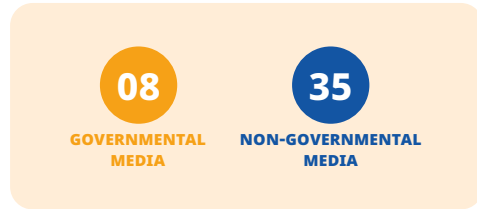
Figure 2: Classification of media pages

Fifth, Classifying the Data:

Steps 1 and 2 resulted in the identification of 1,855 Facebook posts that met the above criteria. Al Hayat then classified those posts and selected those that prompted four or more comments as the final data set. This resulted in the selection of 446 posts, with a total of 11,255 comments.

---

[6] Keywords in Arabic: انتخابات، الانتخابات، مجالس المحافظات، الهيئة المستقلة للانتخابات، راصد، نسبة الاقتراع، صناديق الاقتراع، اعداد المقترعين، نتائج الانتخابات، الفرز، نتائج الفرز

Number of pages from which posts were extracted: 32

Data Sample

Number of posts: 446

Number of comments: 11255
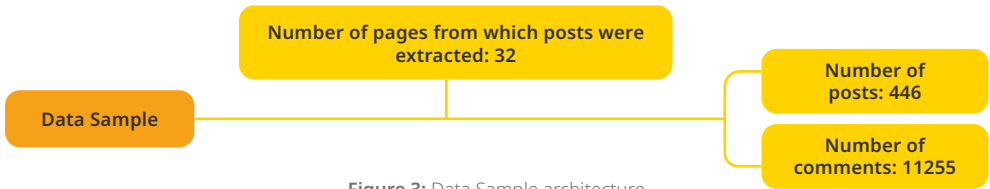
Figure 3: Data Sample architecture

News Site : 16

Media News Company: 8

Radio Station: 5

Topic Newspaper: 4

Broadcasting Media Production: 2

TV Show : 1

Activity General: 1

Figure 4: Categories of pages from which the posts were extracted

23   33   28   1   151   161

Live Video complete   Photos
Native video   Statut
Youtube   Link
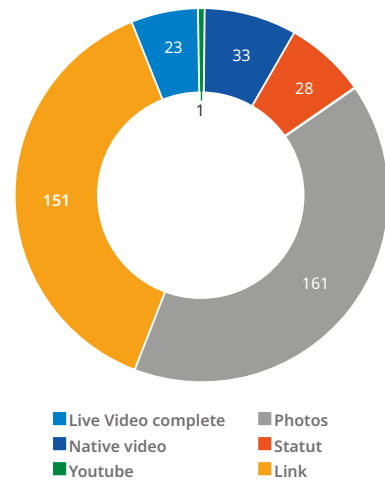
Figure 5: Types of posts (446)

Sixth, Hate speech classification and Intensity scale analysis:

After determining final data set, the SMM team classified the hate speech type and intensity, using the hate speech classification and intensity tables.

1.4. Findings

1.4.1. Data Classification

Comments analysing

The Al-Hayat Centre - RASED team monitored the content of 51 Facebook platforms. From 11,255 comments, the analysis identified 2,630 (23.4 per cent) as containing hate speech, with 64.9 per cent of these comments classified as falling under freedom of expression, and 11.7 per cent classified as "other".

The comments that were classified as a «other» contained advertising or promotional speech that was not related to the post to which they were related.
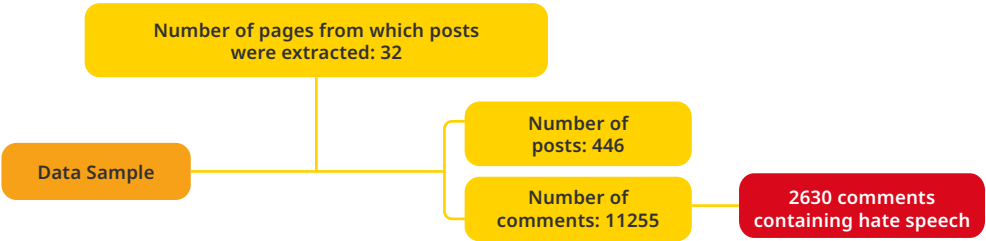
Figure 6: Number of comments containing hate speech

Classifying comments by forms of hate speech

It was found that defamation was the most common form of hate speech, representing 25.78 per cent of the total comments that contained a form of hate speech, followed by denigration, (25.4 per cent), cyberbullying (18.97 per cent), and insults (10.72 per cent). The graph below provides a full breakdown of the types of hate speech by classification.

Figure 7: Classification of hate speech forms

Classifying comments according to their intensity (the strength of their impact)

Classification of Hate Speech comments on Facebook during the municipal and provincial elections 2022

2630 comments

Death • Violence • Demonizing and Dehumanizing • Building a negative character • Negative Actions • Disagreements

678 Defamation — 6.7% 11% 82% 0.3% 0% 0%

22 Harassment — 0% 0% 91% 4.5% 4.5% 0%

668 Denigration — 2.1% 7.5% 85% 4% 0.4% 0%

162 Disinformation — 2% 8% 90% 0% 0% 0%

41 Rumors — 10% 10% 78% 2% 0% 0%

4 Violence — 0% 100% 0% 0% 0% 0%

499 Bullying — 0% 0% 1% 88% 8% 3%

172 Discrimination — 0% 1% 1% 90% 5% 3%

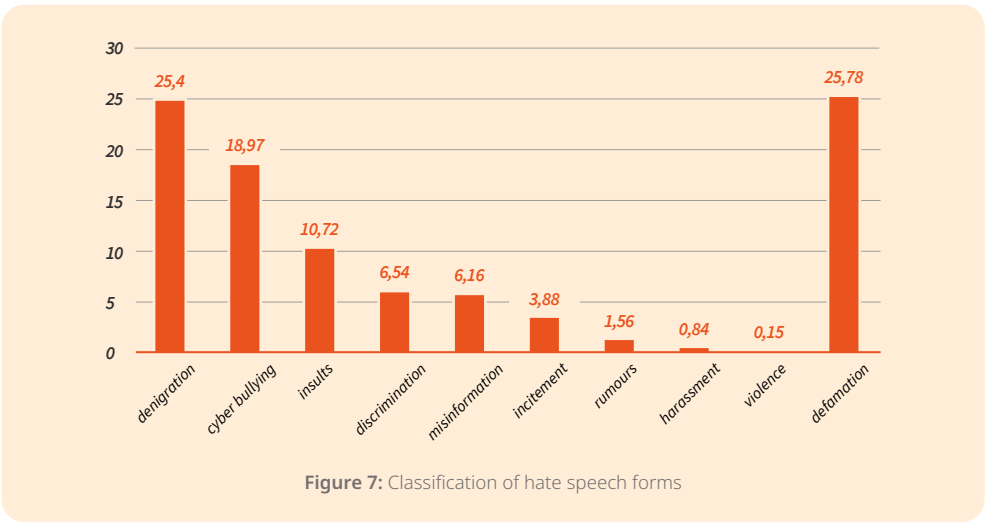**102 Incitement** — 0% 48% 2% 19% 26% 5%

282 Insulting — 0% 0% 10% 71% 14% 5%

1.4.2. Examples of hate speech on Facebook: Behaviours and phenomena

Defamation

Defamation was the most common form of hate speech, accounting for 25.7 per cent of the total comments that contained hate speech. It was found that the monitored behaviours crystallised around defamation of the electoral process and its integrity. For example, one of the comments read «What is this time, the 20s or 50s JOD (Buying Votes)». The monitoring process also identified behaviours related to defamation against both men and women candidates. For

example, one of the comments was *"Because they are all liars, but they all like MPs".* Another said, *"In order to elect corrupt people, boycott".*

Denigration

Almost as common as defamation was denigration, or the phenomenon of belittling others, whether they were individuals, entities or institutions. The share of cases of denigration as a form of hate speech was high, making up 25.4 per cent of all comments that included hate speech.

These behaviours were linked to belittling others, by describing them in inappropriate or insulting terms, for example, one of the comments read *"the country has enough bandits, seriously we are seeing them in the morning and in the night"* [7] , while another read *"Seeing your face cuts off my appetite"* [8].

1.5. Recommendations

- Clearly define and regulate political campaigning mechanisms on social media platforms in the legislation related to elections.

- Include explicit penalties/ sanctions in executive directives and legislation for those who commit bullying, electronic violence, electoral violence or any form of hate speech, with clear written definitions of these acts

- Have the IEC form observatory teams for social media platforms, in association with civil society and media representatives, with the main mission of pursuing and tracing hate speech.

- Legislation should introduce a short and simplified procedure against those who practice hate speech and incitement to hatred or violence, to dissuade the public from engaging in these dangerous behaviours. This procedure should include preventive measures, such as closing pages/accounts or suspending them, with reasonable notice, if they are found to disseminate hate speech.

- Collaborate with social media platforms available in Jordan to develop policies that tackle hate speech and false and misleading content. These policies should be adapted to local legislation and the principle of gradation. For example, in the case of the removal of posts or comments that contain hate speech, the account holder should be informed of the fact that they posted content identified as hate speech and warned. Their account can then be blocked if the behaviour continues.

- The government should include information on hate speech and disinformation monitoring

---

[7] Loosely translated from the comment: هي ناقصة البلد سرسرية جد بكفي أشكالكم الي نتصبح ونمسي فيه

[8] Loosely translated from the comment: شوفة وجهك أكبر مبرر لأنها بتسد النفس

mechanisms in the curriculum of media and journalism colleges, as well as build an anti-hate speech strategy by educating student bodies on hate speech and methods to counter it, and on how to contribute to raising public awareness about all forms of hate speech.

- Laws to deal with misinformation and false news need to be introduced or updated, and should work towards providing a safe online space, free from media and other disinformation.

- The relevant authorities should promote initiatives to counter misinformation and false news that focus on social media platforms. These initiatives should work on social media education and fact-checking. Jordan does already have the official platform "Haggak Teraf" ("Your Right to Know) and should work to see other initiatives (private/public) launched.

- Promote the introduction of programmes, activities, and promotions by CSOs to counter hate speech, working both on raising public awareness on the matter and carrying out related research and studies. They can also launch communications campaigns to educate the public on the legal boundaries of freedom of speech.

2. Lebanon political landscape ahead of elections: Analysis of disinformation, hate speech, rumours, and propaganda

2.1. About the Maharat Foundation

Founded in 2006, Maharat is a women-led, Beirut-based organisation, working as a catalyst in defending and advancing the development of democratic societies governed by the values of freedom of expression and respect for human rights.

Maharat has been leading the media law reform in Lebanon and acting as a watchdog organisation monitoring the situation of freedom of expression, media freedom, the safety of journalists and the free flow of information in the country and in the broader MENA Region.

This report describes Maharat's work between February and March 2022 to follow up on the May 2022 parliamentary elections, through different activities that contribute to countering misinformation and disinformation, increasing factchecking opportunities and fostering voter's critical thinking to help them make informed choices. These activities included analysing media and social media discourses and monitoring campaigns in terms of disinformation,

propaganda, hate speech and violence against women; the production of position papers and resources on elections and the media; training journalists and alternative media platforms on the professional coverage of the elections; and factchecking. The analysis of the election and post-election periods will feature in the next regional report.

2.2. Local context

2.2.1. The media Landscape

Lebanon was ranked by Freedom House as the most democratic country in the Arab region in terms of freedom of the press and freedom of expression, and was classified as partly free in the Freedom House report for 2022. Political pressure is stronger than ever, however, and Lebanon's own scoring is regressing, as it was ranked 130 out of 180 countries in the Reporters Without Borders World Press Freedom Index in 2022, compared to 107th place in 2021. Although the Lebanese constitution guarantees pluralism and protection of freedom of expression, journalists and activists are subjected to regular harassment. An increasing number of journalists and activists are being prosecuted and called in for interrogation on vague criminal defamation charges brought by the public prosecutor on behalf of powerful political, financial and religious figures, who have increasingly used these laws to retaliate against and silence criticism, according to Human Rights Watch. Thus, the use of social media by activists exposes

them to arbitrary prosecution on pretexts related to harming civil peace or blasphemy, offending political and religious symbols, or disturbing Lebanon's relations with friendly countries. While journalists are legally protected in defamation cases, this protection applies only if they prove the veracity of the claims being made in their materials.

Across this context, well-known militant political groups seek to manipulate the democratic discourse on social media through what is known in the Lebanese context as "electronic armies", aiming to spread political propaganda and disinformation, as well as launching violent campaigns against political opponents.

Additionally, with every crisis in Lebanon, the information disorder becomes more visible. Many Facebook groups and pages affiliated to the traditional political parties have emerged since the October 17th revolution in 2019. A documentary produced by Maharat analysed the spread of rumours and disinformation at that time, linking it to the context and analysing its impact on the path of events. Months ahead the 2022 elections, social media campaigns started, some of these aimed at demonising protesters and actors for change. In other cases, disinformation aimed at criticising traditional political parties, allegedly demonstrating contradictions in their platforms, such as the campaign led against the Hezbollah member of parliament (MP) Mohamad Raad, in which multiple pages shared the same meme accusing him of hypocrisy while inaugurating a project funded by USAID.

Figure 8: Example of disinformation in a "meme" format that was circulated by multiple pages accusing Hezbollah member of the parliament of inaugurating a project funded by USAID.

Disinformation trends have been seen on the parts of all political parties, each targeting their political opponents.

Figure 9: An example of disinformation that was circulated by traditional political militant pages accusing opponent parties of financing electronic platforms to manipulate the currency exchange rate.

2.2.2. The political context ahead of the parliamentary elections

The preparations for the parliamentary elections began amidst a severe economic and financial crisis. as well as political tensions between those in power over several issues related to the timing of the elections, the participation of expatriates in the local elections, and the implementation of reforms related to the voting process, including the adoption of magnetic cards and "Central Mega Centre" polling stations, which will allow voters to vote in the place where they live if this is different from where they are registered on the voting list.

Additionally, political divisions and security tensions increased against

the background of the Beirut port explosion investigation, which led to the disruption of the government's work and triggered security incidents resulting in deaths and injuries in the Tayyoune area in the southern suburb of Beirut, near the site of a protest against judge Tarek Bitar, who is heading the investigation of the explosion.

Despite this, preparations for the elections were ongoing, the registration of expatriates wishing to vote out of country was completed at the end of 2021, and the candidate registration process opened at the beginning of 2022. The election campaigns actually started at the beginning of February, and the final candidate lists were announced at the beginning of April, less than two months before the vote, scheduled to take place on 15 May for residents, and 6 and 8 May for out-of-country voters.

The record number of candidates and lists showed enthusiasm to participate in the upcoming elections, particularly among youth. The final number of candidates after the formation of the lists was 718, about 60 per cent of whom are from the forces of change and political movements emerging after the 17 October revolution. There were 103 electoral lists, up from 77 in 2018.

According to the results of the monitoring of politicians' discourses conducted by Maharat, political campaigning oscillated between statements questioning the authority's intentions to hold elections and political campaigns around topics that are part of the political conflict, such as Lebanon's neutrality in relation to conflicts in the region, the diplomatic crisis with Saudi Arabia and the Gulf states as a result of a political statement by the Minister of Information Georges Kordahi on what he described as an "absurd war" in Yemen, and Iranian interference in internal Lebanese affairs, through Hezbollah.

Furthermore, some political actors, including Hezbollah leader Hassan Nasrallah, repeatedly accused the new political actors or those belonging to civil society and emerging political movements of receiving financial support from embassies, specifically the United States embassy, in the service of foreign projects, implying that the goals of these alleged actors of change were in line with Israeli interests, which is considered a dangerous discourse that involves provocation and threat.

As for the discourse of politicians related to the economic crisis, this was mainly centred on state corruption, the recovery of stolen assets and forensic audits. These topics were present in the discourses of most of politicians as part of accusations against political opponents, rather than as part of providing solutions or relevant electoral programmes.

2.2.3. Objectives and methodology of the monitoring programme

2.2.3.1. Objectives

Social media has a key role in political communication during election campaigns, specifically that it provides unlimited interactive means for political influence, propaganda and for direct misleading communication between the candidates and the voters. By extension, the use of such social platforms facilitates the spread of hate speech, false news and misleading information, and can even influence voting patterns in elections through the "services" provided by these platforms, including promoting unverified information, suppressing and intimidating voters, and deceiving voters through false or misleading affiliations.

The monitoring programme launched by Maharat in cooperation with DRI during the campaigning for the May 15, 2022, parliamentary elections monitored hate speech and misleading political discourse intended to influence voters on social media. The monitoring focused on politicians' statements and candidates' personal pages, on false news and rumours that spread through the pages of political influencers or those run by partisan supporters, and on campaigns of manipulation on social media during the campaign period.

2.2.3.2. Monitoring period

The period of monitoring and in-depth tracking for the purposes of issuing the first report ran from the 1 February to 31 March 2022.

2.2.3.3. Monitoring sample (retrieved from Facebook and Twitter)

The monitoring covered 155 actors based on Twitter, as the primary monitored platform, and on Facebook, as the secondary unit for monitoring (only for active users on this platform). These comprised:

- Fifteen first-ranked politicians from among the elite in their parties and active on social media.

- One-hundred-and-seven candidates for the parliamentary elections active on social media and representing the various traditional and «revolutionary» party movements;

- Thirty-three men and women journalists and influencers affiliated with political parties and coalitions, and specifically those with a high number of followers, who have high engagement and interactions with the public and reflect the views of the various political actors: and

- Forty-seven partisans' pages and Facebook groups that have the direct endorsement of a political group, daily posting, and interaction with posts, as well as anonymous digital activity in favour of political forces.

24
ACCOUNTS MONITORED ARE FEMALE

131
ACCOUNTS MONITORED ARE MALE

2.2.3.4. Adapting the Toolkit to the specific monitoring strategy and methodology

Maharat's monitoring team has been using a wide range of tracking, monitoring and analysis tools to track and counter manipulation campaigns.

The most prominent tools used and their purpose:

1. Maharat Tracker software for daily tracking of the political discourse of politicians, candidates, and influencers of the targeted selected sample. It works on gathering data based on a combination of accounts and keywords from Twitter, facilitating subsequent manual classification and analysis. Maharat Tracker's infrastructure allows for predictions at a later stage.

2. The CrowdTangle program facilitates the process of monitoring the pages and groups of electronic armies on Facebook, by preparing specific lists that allow for tracking political propaganda and disinformation discourses, specifically in manipulated images and memes, as well as identifying the main sources of manipulation and misinformation campaigns.

3. The TweetDeck feature allows for tracking the source of campaigns on Twitter when they are launched, as well as the most prominent hashtags used.

Open monitoring and tracking tools:

- VICINITAS, which provides big data collection on Twitter related to specific campaigns, based on the use of campaign hashtags, keywords or account addresses.

- Hoaxy-, which allows for the identification of the structure, interconnectedness and influencers of networks.

- Fake account verification tools, such as Botometer and Bot Sentinel, to detect bot behaviours.

2.2.4. Monitoring streams

2.2.4.1. First stream: Monitoring the accounts of politicians and candidates in the parliamentary elections

This involved monitoring the accounts and pages of political actors, including candidates and politicians, as well as the pages and groups of the "electronic armies", through daily tracking of all statements and publications, which were classified according to the monitoring and analysis mechanism below:

A- Daily monitoring of the misleading and violent political propaganda speech on any of the pages mentioned above and on online accounts.

B- Classifying the misleading discourse used: The following classification was adopted for this analysis:

- **Disinformation:** Fabricated news that has no basis in truth.

- **Misinformation:** News that misleads and distorts the facts, and includes incorrect or misleading information.

- **Rumours:** The circulation of unverified information, that might be true, partially true, or entirely false.

C- Classification of violent speech:

- **Hate speech:** when there is an invitation or direct incitement to violence or discrimination and individual based on religion, gender, race or colour; and

- **Violent speech:** speech intended to provoke general feelings of anger or violence.

2.2.4.2. Second stream: Monitoring campaigns and tracking positions based on the most controversial topics of the day

As election campaigns address different controversial topics, each party seeks to use these topics according to different misleading narratives that serve its propaganda campaign, with the aim of influencing voting tendencies.

DATE	#HASHTAGS	NEW/OLD	PARTY	TARGET	EVENT
24 Feb 11:30	#Civil-Lebanon	New	Free Patriotic Movement (FPM)	Public Opinion	Political campaign launched by FPM as an invitation to the "Civil Lebanon" conference.
6 March 10:34 pm	#Sanyoura-got-off-Beirut	New	Public Opinion	Fouad Sanyoura	Reactions to rumours about Sanyoura's lead and running for elections in Beirut.
7 March 5:26 pm	#ISIS_14_March	New	Hezbollah	Revolution	Reactions against removing the pictures of Qassem Souleimani in the Book forum.

DATE	#HASHTAGS	NEW/OLD	PARTY	TARGET	EVENT
11 March 2:05 pm	#The_lying _revolution	New	Free Patriotic Movement	Revolution	Gebran Bassil short video attacking and accusing the revolution.
24 March 1:41 am	#The_terrorist _Hezbollah	Old	Lebanese Forces	Hezbollah	President Aoun visit to the Vatican and credits given to Hezbollah presence in Lebanon.

Table 3: Most used narratives by political parties to serve their propaganda

2.3. Summary of the monitoring results of the election campaign period on Twitter and Facebook (February – March):

2.3.1. Main findings on political propaganda among politicians, candidates, influencers and partisan's electronic groups and pages, known in the Lebanese context as «electronic armies»

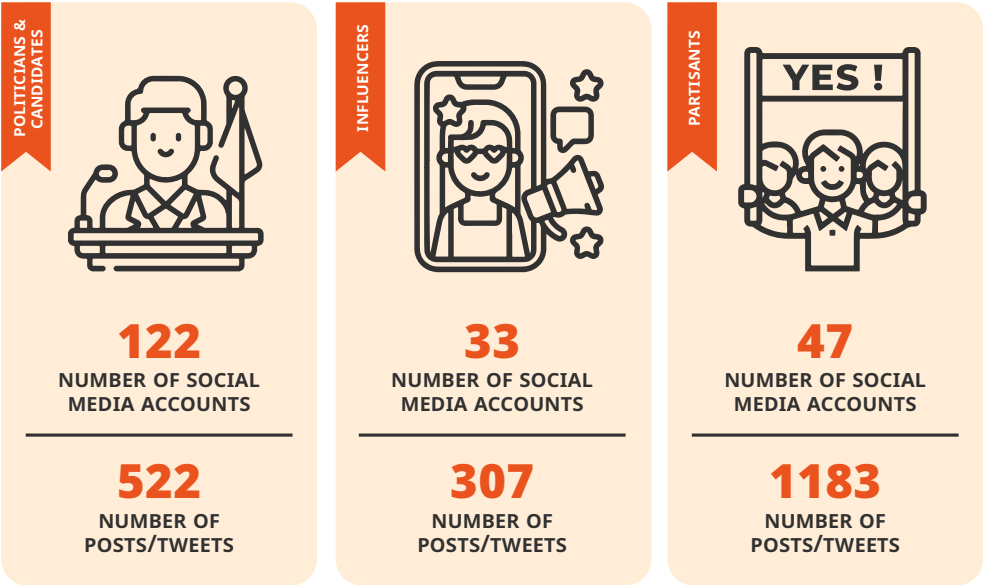The following data indicates the number of tweets and posts monitored during the months of February and March:

POLITICIANS & CANDIDATES	INFLUENCERS	PARTISANTS
122 NUMBER OF SOCIAL MEDIA ACCOUNTS	**33** NUMBER OF SOCIAL MEDIA ACCOUNTS	**47** NUMBER OF SOCIAL MEDIA ACCOUNTS
522 NUMBER OF POSTS/TWEETS	**307** NUMBER OF POSTS/TWEETS	**1183** NUMBER OF POSTS/TWEETS

Table 4 : Sample of data collected during February and March 2022

A. Political campaigning Speech on social media:

- The total number of tweets and posts published by the 122 candidates and politicians in the monitoring sample during the months of February and March was 522.

- The tweets and posts were mainly of two types: playing on the emotions (257 tweets/posts) and accusations (173 tweets/posts). "Conspiracy theories" were the third-most common (34 tweets/posts). The first two types aimed mainly at defending their own positions and addressing accusations at their opponents. The same was the case for the "conspiracy theory" posts and tweets, as they all aim at triggering partisan emotions. The number of tweets and posts that provided an analysis of events (21) and the electoral programmes (6) was low.
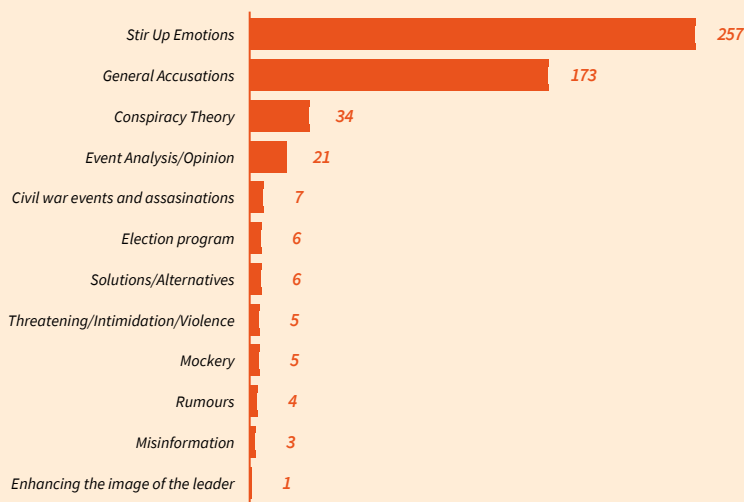
Figure 10: Breakdown of 522 tweets and posts by candidates and politicians

Most topics discussed by the monitored **candidates and politicians** highlighted corruption, while refugees ranked in last place among the topics of interest in their discourses. (Figure 11 below)
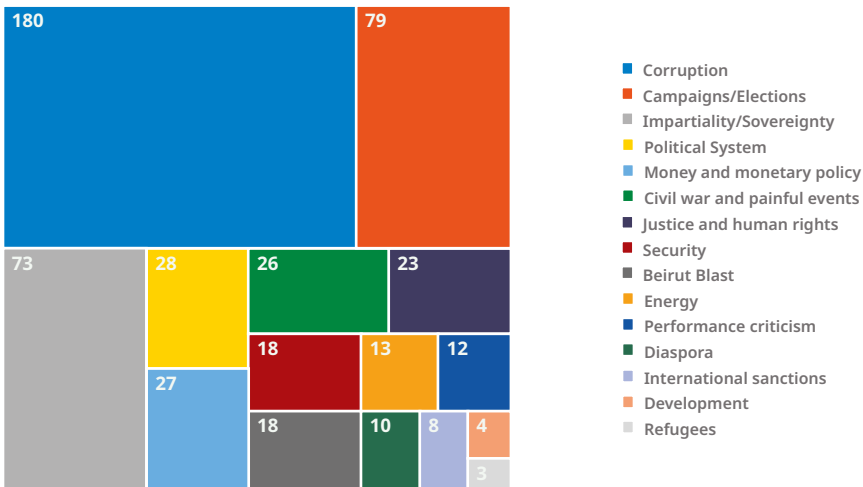
Figure 11: Most Discussed Topics by candidates and politicians

The graph below shows the breakdown of the type of political communication used by candidates during the election campaign, based on a sample of 358 tweets and posts in the months of February and March 2022.
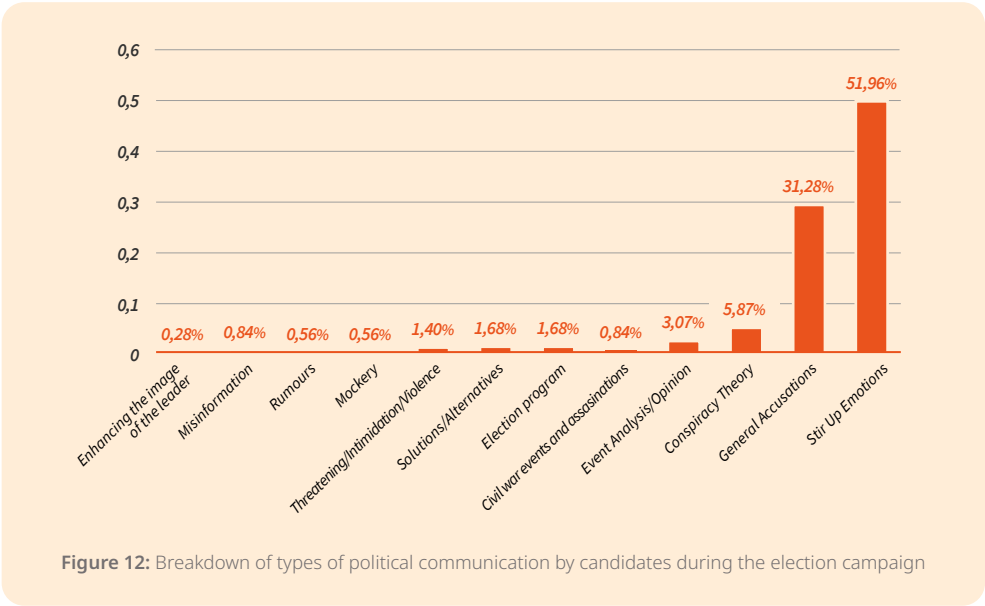

Figure 12: Breakdown of types of political communication by candidates during the election campaign

Early observations from monitoring the activities of candidates and politicians on the Twitter and Facebook platforms during February and March are that their speech was primarily in the form of propaganda and promotional speech, aimed mostly at stirring up negative feelings and adopted general projections and accusations by more than 80 per cent. This applies to traditional political forces as well as to emerging political movements, even though the emerging powers were expected to promote a reform project and/or rational criticism of the forces that they seek to replace. The contents and tweets of the emerging powers focusing on their electoral programmes or proposing solutions to current problems were 3 and 2.5 per cent, respectively.

It is worth noting that the content of politicians' tweets and posts was almost completely devoid of solutions or reform suggestions, as well as any electoral programmes, resulting in a political discourse focused on inciting emotions rather than proposing rational solutions to citizens' problems.

Despite the fact that topics related to corruption and election campaigns

were among the most common in the tweets and posts of the political actors as shown in table 3, these political discourses of the candidates and actors was mainly focused on making accusations and playing on emotions, with a share of over 80 per cent, as well as misleading through the promotion of conspiracy theories, rather than building on facts and evidence or electoral programmes, as per tables 2 and 4.

This applies to both traditional political forces and to emerging political movements, although the emerging powers were expected to offer reform projects and/or rational criticisms of the forces that they seek to replace. The contents and tweets of the emerging powers focusing on electoral programmes or proposing solutions were just 3.3 per cent and 2.5 per cent, respectively, of all examples. These figures are extremely low for electoral preparation and campaigning.

A full 94.2 per cent of tweets and posts can be described as *"populist"*, aiming to influence voters away from topics such as needed reforms, and avoiding topics such as the needs of the society or voters' expectations.
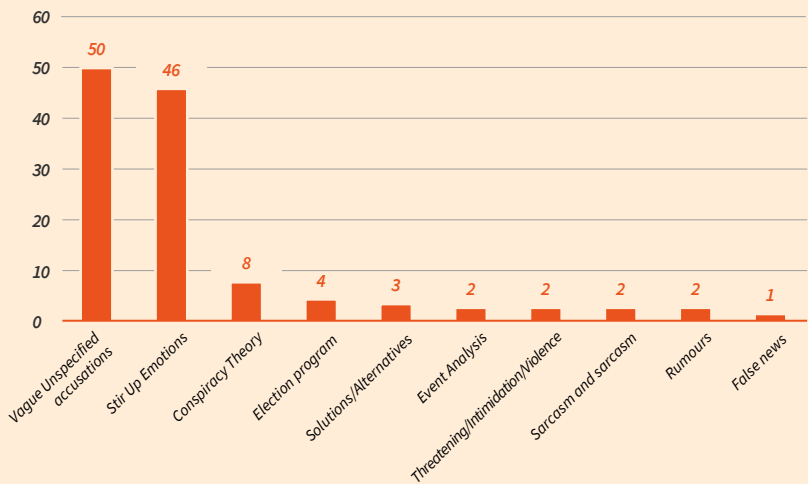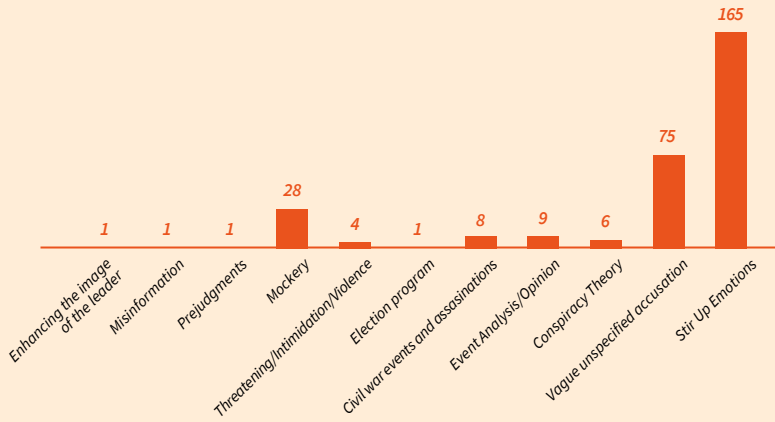
Figure 13: Break down[9] of the type of political propaganda used by candidates and politicians from emerging political movements during election campaigns

B. Influencers and partisans' participation in political communication during the election campaign:

Apart from the political actors and candidates, the report monitored the discourses of 2 additional types of actors on social media: influencers with different political affiliations, and pages directly endorsing specific political groups, whether traditional or emerging.

- Three-hundred-and-seven tweets and posts for the influencers were monitored, 255 of which were by influencers from traditional forces and 52 by those from emerging forces.

---

[9] The sample is from 120 monitored tweets and posts for the months of February and March 2022

Figure 14: Break down[10] of the type of political propaganda used by influencers from emerging political movements during the election campaign.

The most-discussed topics by the monitored influencers focused more on corruption than on refugee-related themes. (Figure 15 below)
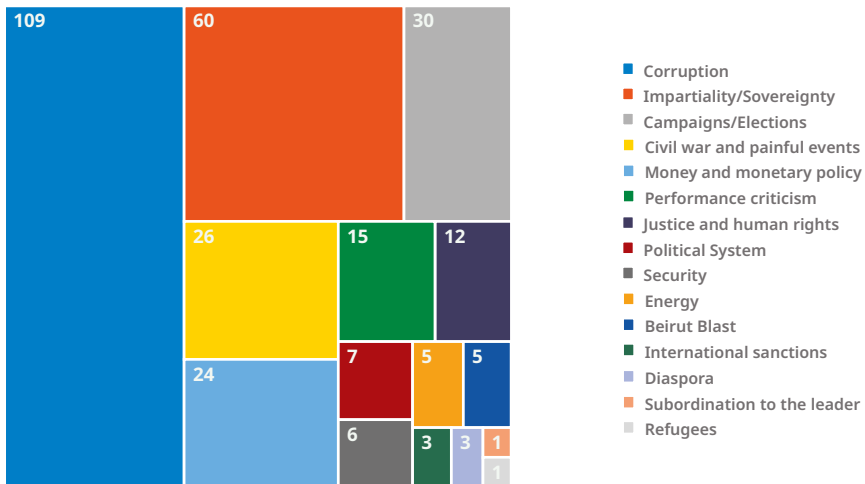
Figure 15: Most Discussed Topics by influencers

[10] The sample is from 307 monitored tweets and posts for the months of February and March 2022

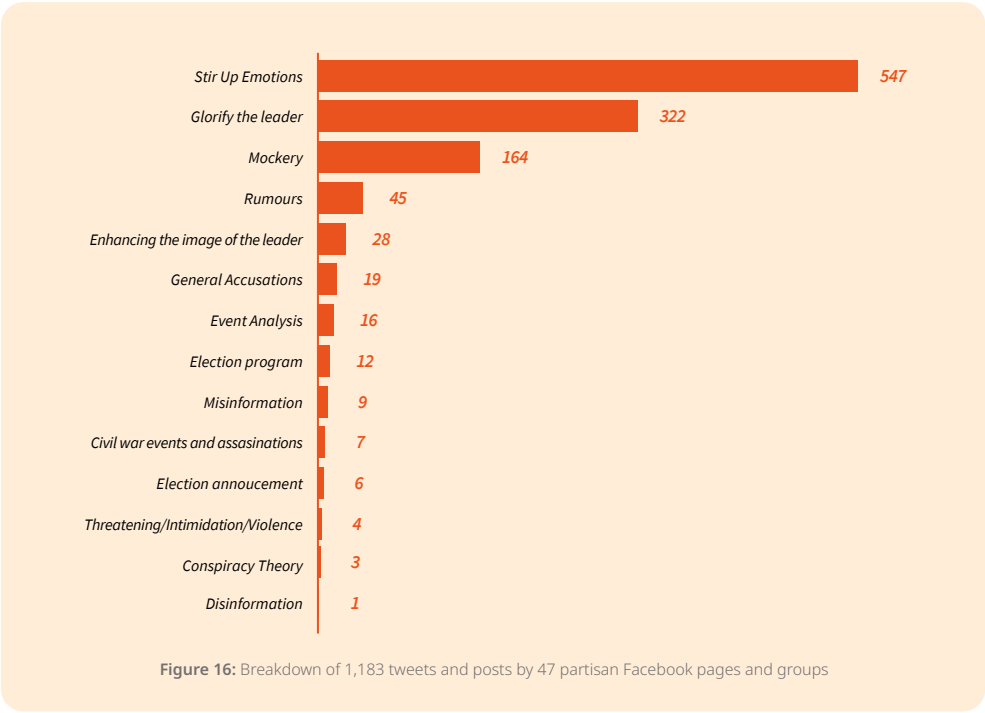Examining the political communication of monitored influencers, the type of campaigning and propaganda was the same among the two types of influencers: those supporting traditional powers and those supporting emerging powers, while the expectations were different.

The three main types of discourse among both types of influencers were playing on emotions, accusations and sarcasm. It is notable that the percentage of playing on emotions was higher in the discourse of the emerging groups, where it was 57 per cent, versus 52 per cent for those associated with traditional powers.

- There were 1,183 posts monitored on Facebook pages, issued from 47 accounts, with content promoting certain political parties.

Stir Up Emotions	547
Glorify the leader	322
Mockery	164
Rumours	45
Enhancing the image of the leader	28
General Accusations	19
Event Analysis	16
Election program	12
Misinformation	9
Civil war events and assasinations	7
Election annoucement	6
Threatening/Intimidation/Violence	4
Conspiracy Theory	3
Disinformation	1

Figure 16: Breakdown of 1,183 tweets and posts by 47 partisan Facebook pages and groups

Examining the political communications of monitored "partisan groups and pages" on Facebook, it is clear that these groups either aimed at defending their leaders and attempting to enhance their reputations, or at spreading rumours against their political opponents. The identity of the owners of these pages remains unknown, however, and can only be speculated upon based on their content.

2.3.1.1. Main findings for campaigns to manipulate conversations on social media.

The manipulated news disseminated through these accounts accounted for 4.64 per cent of the 1,183 publications, the majority of which were rumours.
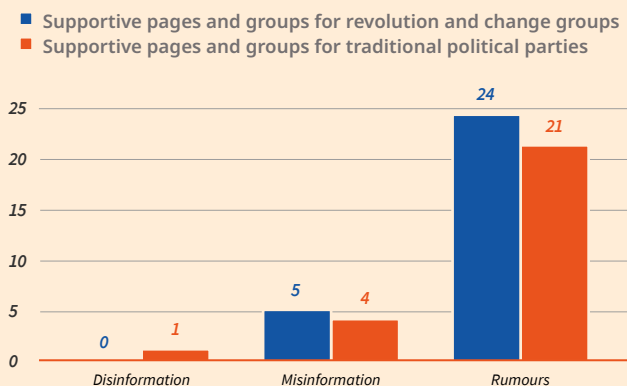

Figure 17: Manipulated content by electronic armies on Facebook pages and groups from 1,183 posts

Manipulation campaigns were not limited to misleading news and propaganda but were also used in campaigns aimed at defaming and discrediting political opponents.

These campaigns do not always appear to be part of a trend on social media, as the Maharat team monitored an individual programmed manipulation campaign against a journalist and political activist close to one of the political parties.

The most prominent characteristics of the manipulation campaign mentioned above were:

- This campaign was launched on February 17.

- Fifteen accounts used in this campaign were created in December 2021 and became active on 11 January 2020.

- All the accounts had zero followers both before and throughout the campaign.

- Pictures and names of public figures were used for profile accounts to make them seem real.

- The usernames for the profiles include non-sequential and random numbers.

- The page owner used a defamatory hashtag, indicating that this journalist is promoting politicians who pay them for this.

- The same tweet was published at the same time by several accounts, and in the same order.

- Through Botometer, it was determined that all these accounts are bots.

2.3.1.2. Main findings for hate speech campaigns during the elections on social media platforms

Hate speech on social media platforms

The following phenomena were tracked and examined in the monitoring process led by the monitoring team during the period from 1 February 1 to March 31 to identify hate speech disseminated during the election campaign period on the Facebook and Twitter platforms:

Facebook groups affiliated to Lebanese parties

By monitoring several pages and groups supporting Lebanese political parties, a number of non-coordinated campaigns in which hate speech was used against their opponents were found and analysed. Among the hate messages that were monitored were the following:

Hate speech message disseminated in same time by a partisan of Progressive Socialist Party on several Facebook groups saying:

«This crowdn which was on the 19th of March, will hammer down a nail into your coffins … We will not be merciful anymore.»

#Progressive_Socialist_Forum_Aley
#Media_affairs

Figure 18: Examples of screenshots of hate speech messages by Future Movement partisans

Another hate speech message spread by future movement partisans:

«We do not have a president; we have an Iranian intelligence boy. We must summon you to Martyr's Square and achieve justice for you by hanging.»

Hate Speech: «Dalia Ahmed» case

On 13 January 2022, Dalia Ahmad, the media host of Al Jadeed TV, was subjected to a racist harassment campaign on social media, based on her colour and Sudanese origin. The racist campaign was launched due to her criticism of Lebanese politicians, including President Michel Aoun and Hezbollah General Secretary Hassan Nasrallah, whom she called "*crocodiles*", on her satirical political television programme.

The hashtags #Dalia_Ahmed and #itahseen_nasel_kaleb ("improving dog strain") were trending for more than 2 days. The campaign was renewed on 24 February, after Ahmad sarcastically criticised Nasrallah, who she said had become distracted by moving wars in the region, at the expense of Lebanon's interests, and the speaker

and head of the Amal movement, Nabih Berri, whom she accused of corruption. She also criticised the 2 Shia symbols in Lebanon,

The following findings have been monitored:

- The second campaign on Ahmad started on 24 February by a user who wrote on Twitter: "*The black wicked #dalia_ahmad insisting boldly to insult our symbols*".

- The campaign targeting the television host is an example of a racist hate campaign on social media, to which militants from some Lebanese political parties contributed by disseminating the tweet widely on Twitter, using the hashtags ‎#داليا_احمد‎ and ‎#تحسين_نسل_الكلب‎

- This campaign had a gender dimension, consisting of violent discourse against women, aimed at harassing, threatening and silencing Ahmad.

- The campaign included multiple forms of violence against women:

    ○ **Racial and psychological violence**: based on her appearance and skin colour; and

    ○ **Sexual violence:** calling her an "illegitimate girl", an epithet for prostitutes.

1. Among the violent expressions and terms used in the campaign against Ahmed included:

- *"Ethiopian black";*

- *"black dog";*

- *"Sudanese bastard";*

- *"Dalia_Ahmed is a prostitute"*

- *"Pimp and a thug's wife, who works for a sucker: and*

- *"The damned black witch".*

Another campaign was launched in support of her on social media, denouncing racist expressions and supporting her right to expression.

Other activists [11] appeared to consider harassing Ahmad their "duty", in order to incite the campaign against her.

---

[11] The tweets, respectively, say:

- «The black-haired #Dalia_Ahmed is brazenly insisting on insulting our symbols.»
- #Dalia Ahmed, the bad reputed girl from Sudan, is daring again to insult al Sayed. A Sudanese whore
- #Dalia Ahmed, a title of prostitution in Lebanon.
- Bullying #Dalia Ahmed is a duty.

2.4. Conclusion

While campaign periods in democratic elections are considered opportunities for candidates to create platforms where they share their programmes, visions and solutions to the crises their country is facing, the social media monitoring conducted by Maharat showed that of the political discourses in February and March, ahead of the 15 May elections, consisted mainly of playing on voters' emotions and sharing unfounded accusations. This was the case in over 80 per cent of the materials monitored.

This applies to the activities of both traditional and emerging powers. Whereas the latter were expected to promote reformist programmes and provide evidence-based criticism, only 3 per cent of their discourses focused on their electoral programmes and only 2.5 per cent focused on providing solutions to the country's crises.

Topics and discourse that are considered emotional stirring for the Lebanese people occupied most of the social media space and focused around 4 main topics: corruption, elections, sovereignty, and neutrality. These are mainly topics of conflict among political actors although they are not fact-based, and the discourse around them consisted mainly of making accusations against political opponents. Eight per cent of the materials identified were promoting conspiracy theories, with their parties portrayed as the victims.

The partisans of political groups took the same approach, by adopting their leaders' discourses and defending their positions, while spreading rumours and disinformation. Rumours and disinformation were practically absent from the discourses of politicians. Many examples of this were identified during the reporting period, including the disinformation campaign against Hezbollah MP Mohamad Raad and the hate campaign against journalist Dalia Ahmad, for her statements against Hezbollah.

2.5. Recommendations

- Enhance the engagement between CSOs monitoring social media and social media platforms, to combat disinformation and hate speech more effectively.

- Increase the capacities of social media users, candidates and civil society members to detect hate speech on social media platforms, and to report accordingly.

- Increase digital media information literacy skills among users across Lebanon, to combat disinformation and hate speech.

- Further build the capacities of journalists and activists on factchecking, to increase such initiatives across the country.

- Engage with media and public actors to increase voter education materials available.

- Launch initiatives to foster a culture of accountability, in which candidates are held accountable for their electoral programmes and promises.

3. Tunisia in the aftermath of the 25 July 2021: New disinformation streams

3.1. Réseau Mourakiboun

"Réseau Mourakiboun", commonly known as "Mourakiboun", is a Tunisian association created in 2011, with the aim of reinforcing and supporting the post-revolution democratic transition in the country. Since its creation, Mourakiboun has focused its efforts on election observation. The professionalism, impartiality and hard work of Mourakiboun's 3,000 observers have earned the association great credibility among the Tunisian people, as well as among the authorities, and it is now considered the main player in the observation of electoral processes.

In 2014, Mourakiboun launched the "RASD" project, in partnership with DRI and the Open Society Foundations. "RASD" represented a turning point in Mourakiboun's history, as it took election observation one step further, by being the first national initiative to observe candidates' activities on social media networks. During the 2014 presidential election, "RASD" observed the candidates' activities on Facebook, Twitter, and YouTube, and produced a report with statistics and analysis of the performance of the candidates on these social media networks.

In addition to its reputation in election observation, Mourakiboun has been acknowledged for its work in awareness-raising and advocacy. It has successfully led various awareness-raising campaigns focusing on decentralisation principles, the roles and prerogatives of local authorities', and women and youth participation in political life at the local level. Mourakiboun has led advocacy campaigns and actions related to the electoral process, aiming to improve the organisation of elections and to guarantee their transparency. Advocacy campaigns were carried out in cooperation particularly with the Independent Higher Authority for Elections (ISIE), political parties, deputies from the parliament, media agencies and CSOs.

3.2. The Institute of Press and Information Sciences

The Institute of Press and Information Sciences (IPSI) is one of the oldest academic institutions in Tunisia and, as the only public university teaching communication and information sciences, it has built a strong reputation over the years. Its former students currently work in various public and private media throughout the country, as well as abroad for international media.

IPSI's contribution to research in information and communication

sciences is no less important than the training of media professionals. Thanks to its scientific journal, international symposiums, research units and the research carried out by its teachers and students, it has acquired great international influence.

3.3. The "LAB'TRACK" project

The Tracking Laboratory, or *"LAB'TRACK"*, project joins the respective efforts of Mourakiboun and IPSI to dig deep and understand the phenomenon of disinformation and the manipulation of online political content, in order to enhance a strong and sustainable democracy. The focus of the project is to explore the phenomenon of online disinformation, with a focus on Facebook, and identify social media platform users' behaviour, as well as mobilise different stakeholders to engage in a dialogue to discuss the risks and threats online disinformation pose for society and democracy, and to work with political decision-makers to establish public policies to fight online disinformation.

3.4. The Country context

On 14 January 2011, Tunisia witnessed an important day in its modern history, when President Mohammed Zin El Abidin Ben Ali fled the country after a series of demonstrations across Tunisia. Under the Ben Ali regime, the Tunisian authorities had for a long time exercised a virtual embargo on Tunisians' rights to

information, as the media was under the control of the president and the government and the internet network was censored. But, thanks to many digital rights activists, social media platforms, and especially Facebook, played an enlightening role in the circulation of information in the months leading up to the popular uprising.

3.4.1. The local context and how the "LAB'TRACK" project will address it

The online disinformation phenomenon has threatened Tunisian social media users for several years already, but it is only recently that some initiatives have been launched to study and dig deeper into this. These initiatives have shed light on online disinformation, with a focus on disinformation on Facebook, as it is the platform used most by Tunisian internet users. The COVID-19 pandemic has been accompanied by a large spread of disinformation on social media platforms and, again, especially Facebook, and this led to national initiatives to look at this phenomenon and inform the public, which needed concrete and clear medical information to survive the pandemic. With the lessening of the pandemic emergency, only a few initiatives continue to work on disinformation related to COVID-19, and their focus has shifted mainly to political issues.

The *"LAB'TRACK"* project aims to increase public awareness of false

news and to aid in avoiding it, by familiarising people with the machinery and agendas behind the phenomenon. This understanding comes in different depths, depending on the targeted group itself.

Tunisian citizens are the main target beneficiaries of the project, with a focus on human rights activists, journalists (those working in the profession, as well as professors and students at IPSI) and politicians

3.4.2. What to monitor and why?

The social and political content on *"Tunisian Facebook"* (the local Tunisian virtual social media sphere of Tunisian accounts, pages and groups) has been the same since 2011. The monitoring reports of the 2019 Tunisian presidential elections demonstrated how Facebook was used to influence and impact the choices of some voters. If not positively convincing them for whom to vote, it nonetheless was used convince voters not to vote for certain candidates. So, the main materials for the laboratory will be the posts the *"LAB'TRACK"* team suspect to be "fake news" as part of viral actual campaigns.

The *"LAB'TRACK"* team has identified several themes that will be the focus of its monitoring missions, while remaining ready to include any others that might become relevant as a result of any surprising event. *"LAB'TRACK"* focuses on "Disinformation and political manipulation on social media platforms after July 25th", as the 25 July 2022 constitutional referendum is

the new highlight in Tunisian modern history, following the 14 January 2011 referendum. The monitoring covers six thematic areas, including: The national consultation, the dissolution of the parliament, Women in politics The PM "Najla Bouden" as a case study, the referendum, the economic crisis and organising the elections.

The anticipated output of the monitoring of these topics is to answer theoretical questions about the existence and prevalence of political violence against women, political manipulation, manipulation and misinformation of the referendum process, misinformation and manipulation of the economic crisis, and misinformation and manipulation of the election process. *"LAB'TRACK"* will produce six observation briefs that will assess behaviour related to this phenomenon, and one final report that explores the disinformation phenomenon in Tunisia, its roots, and its impact.

3.4.3. Monitoring tools:

The main tool to be used is CrowdTangle.

https://www.theguardian.com/ books/2019/nov/22/factitious- taradiddle-dictionary-real-history-fake- news

4. Sudan: Online hate speech in the aftermath of the coup of 25 October

4.1. Context

Over six months following the military coup in Sudan, announced on 25 October 2021during a speech by the Commander-in-Chief of the Sudanese Armed Forces, General Abdel Fattah Al-Burhan, who is also the Chairman of the Sovereignty Council, hopes that a return to a civilian-led transitional period that can rid the country from the burdens of civil war, military rule and a woeful economy have become increasingly faint. The 25 November agreement signed between General Al-Burhan and Prime Minister Abdullah Hamdok, according to which Hamdok returned to the exercise of his duties, did not last long, and the prime minister's subsequent resignation, on 2 January 2022, has left the military in full control of the country's affairs, further complicating the situation and diminishing the chances of a return to the civilian-led, democratic rule of the transitional period.

Over 2 billion dollars of international financing in support of Sudan's ailing economy continues to be frozen, and progress towards relief on the country's eligible debt has stalled,

with creditors reluctant to finalise agreements in the wake of the coup. The economy continues to deteriorate, with triple-digit inflation and a decrease in the value of the Sudanese pound leading to ever-increasing prices for basic goods and services. The conflict in Ukraine and its global economic repercussions suggest a more dismal outlook for the economy going forward.

Street protests and demonstrations have become almost daily occurrences and have been met with excessive use of force on the part of the military. The escalating harassment of protesters, including abductions, detentions, and deaths, has undermined mediation and dialogue efforts championed by international actors such as the United Nations International Transition Mission in Sudan (UNITAMS), as well as other nationally driven initiatives. The political impasse continues unabated, with political actors and forces unable to reach agreement on the way forward.

Within this context, social media is having an enormous effect on shaping the public's understanding of the situation and influencing the positions of key political actors and their political parties regarding the political impasse and the way forward. Social media, as noted by SUDIA team during the preparation of the project, is increasingly being seen by many as playing a negative role during this difficult time in the country's trajectory – fuelling fragmentation and polarisation and sowing the seeds of

hate and manipulation. Government efforts to curb the spread of hate speech and dis/misinformation on social media and the internet have, thus far, been limited to decrying the phenomenon in public speeches and enforcing laws and policies that stifle freedom of expression, and that are increasingly being used to intimidate and silence political opposition. For example, on 21 November 2021, the Sudanese Journalists Network issued a statement condemning the decisions by General Al-Burhan to form a legal committee to follow up on all news related to the army and the Sovereignty Council, with the power to criminally prosecute journalists or media outlets, stating that this decision limits freedom of expression and of the press.

Without an adequate understanding of and evidence regarding the proliferation of hate speech and online manipulation on social media, and if these are left unchecked, the political divide and fragmentation will only increase. Actors looking to facilitate dialogue and bridge differences find themselves fighting an uphill battle in an increasingly hostile cyber environment, characterised by cyber-bullying and defamation. Meanwhile, capacity to understand the phenomena and promote actions that might deter its spread and negative impact is extremely limited, with hardly any efforts being made by national actors in this regard.

As part of its programming on promoting democracy and human rights,[12] and in response to this troubling context, SUDIA will undertake monitoring of social media platforms (Facebook and Twitter) to acquire an understanding of the prevalence and proliferation of hate-speech and misinformation that is fuelling the current political impasse and preventing a resumption of a peaceful transition to democracy in the country. Findings from the social media monitoring will be used to (i) develop and pilot innovative actions that counter the phenomenon, and (ii) to advocate for improved legislation, policies and practices that would serve to curb the spread of hate speech and misinformation, while, at the same time, observing human-rights standards and principles.

4.2. Monitoring efforts

The social media monitoring and analysis will seek to understand:

- How hate speech and political polarisation are fuelled by false positions projected by certain actors and entities regarding specific issues.

- How much influence hate speech and misinformation have on political decision-making?

- Whether the conflict between the military and civilian parties

---

[12] SUDIA programming on democracy and human rights promotion is centred around three areas: (i) monitoring and promoting civic space, (ii) enhancing, and supporting the role of civil society and the media in furthering democracy and good governance, and (iii) furthering the participation of marginalised groups of society in political and development processes.

on social media led to the failure to build transitional institutions or the performance of existing institutions at the time; and

• Whether there are indications that widespread hate speech will threaten the election – one of the most important elements of a democratic transition?

The second step is to develop and flesh out a methodology establishing which social media monitoring platforms the team will monitor, the technical tools that will be used, and the finer details of what and whom will be monitored.

Social media monitoring is limited to monitoring only two social media platforms on the internet, these being Facebook and Twitter. These two were selected because they are the most widely used[13] within the Sudanese context.

For each of these platforms, media monitors will monitor the use and prevalence of specific terms and phrases that represent hate speech.

SUDIA benefits from its previous experience in developing a dictionary of hate speech, in addition to holding a round table meeting that gathered several experts to develop a definition of hate speech within the context of the project, considering global definitions.

The scope of the social media monitoring project focuses on specific terms, phrases or other forms of expression intended to humiliate, vilify or insult, and that may be considered abusive or seditious, and that lead to (i) negative mainstreaming (such as describing an entire group in a negative way, as an example, so the women participating in the demonstrations were described as "matlwgaa," which is a description in Sudanese colloquialism that means women who are notorious in terms of morality) and/or (ii) discrimination based on ethnicity, geography or political affiliation, and/or (iii) violence in any form (verbal, physical, emotional) that contributes to political exclusion and/or polarisation.

Using this definition, a workshop gathering media professionals, politicians, representatives of youth and women groups, and jurists was organised by the SUDIA team to review the initial lists of keywords and accounts prepared. The input from the workshop contributed to the finalisation of the lists, establishing 37 terms and phrases that constituted hate speech. Based on the SUDIA team's understanding of the content of hate speech in the Sudanese context, with respect to the scope of the project feedback from the workshop, the terms were grouped into three categories, depending on

[13] The Digital Sudan Statistics website for the year 2020, which reviews the number of internet users and data of those who use the internet, showed the social media (social networking sites and e-commerce), where the report indicated that the population of Sudan, which amounts to 43.33 million people, of whom 13.38 million use the internet, is an increase of 13 per cent over the previous year, while the mobile phone is used by 32.83 million people, representing 76 per cent of the population, and active accounts on social media have reached 1.30million people. Most of them have Facebook accounts

the type of impact they were seen to have. These were:

1. Negative messaging.

2. Discrimination based on ethnicity, geography, political affiliation, or gender; and

3. Violence in any form (verbal, physical or emotional)

The SUDIA team identified several Twitter accounts and Facebook pages that will be monitored for the type, prevalence and use of the hate speech terms and phrases being monitored. The selection of these accounts and pages was based on the consideration of multiple factors, including:

• The type of ownership of the page/account (official, unofficial, influencer, media, etc.);

• The degree of activity on the page or the account.

• The number of page or account followers.

• The level of engagement with page or account posts (comments, shares, mentions, etc.);

• Diversity in terms of political orientation/representation.

• Gender considerations.

• Qualitative considerations; and

• The knowledge and expertise of the SUDIA team

Applying these factors, the SUDIA team came up with 104 accounts that are being monitored – 63 Facebook page accounts and 41 Twitter accounts. This list was validated and finalized during March workshop with stakeholders.

What tools will we use for the monitoring?

Media monitoring tools being used by the SUDIA team include:

• Twitter API - https://developer. twitter.com/en/docs/twitter-api

• Postman API Monitors - https:// www.postman.com/api-monitor/

• JSON Viewer - https://jsonviewer. com/

• JSON-to-Excel converter - https:// conversiontools.io/convert/json-to-excel

• CrowdTangle - https://www. crowdtangle.com/

Special issue: Gender-based violence on Twitter, a case study from the MENA region,

Helmi Noman

This report contains description of violence against women.

This article examines Gender Based Violence (GBV) online, by investigating whether highly visible politically active women from the Arab region are being targeted by devaluating, demeaning, hurtful or intimidating sexual or racial slurs. The study was conducted by Helmi Noman, a social media researcher and analyst, who has examined the digital landscape in the MENA region at the Berkman Klein Center for Internet & Society, Harvard University. It finds that four women of influence on Twitter included in the study have been targeted by offensive content (text and imagery). The content was sparked by the women's engagement in and commentary on contentious political issues. The offensive content was generated by a few speakers, and was amplified by retweets and replies, forming distinct communities on Twitter around the offensive content. A counternarrative against the use of GBV emerges as part of the discourse, yet the offenders use the counter hashtags to disseminate more offensive tweets.

Sample of Arab women of influence on Twitter

The women in the sample were chosen for their high public visibility and engagement in politics. They are:

Loujain Hathloul

a Saudi women's rights activist and a political prisoner.

Dima Sadek

a Lebanese TV news anchor.

Ghada Oueiss

a principal news presenter for Al Jazeera TV.

Tawakkol Karman

a Yemeni Nobel Peace Prize Laureate.

Research Method

- The study searched Twitter historical data for mentions of the names of the sample members or their Twitter usernames, and for a contextually specific lexicon in Arabic. Both the samples of the women and the lexicon were developed by the GBV Lab, in consultation with CSOs in the region. The lexicon consists of gender-specific slurs and racially offensive terms. They are: قحبة OR شرموطة OR قوادة OR منيوكة OR قردة. (Translation: "whore" or "prostitute" or "pimp" or "slut" or "monkey"). These terms are known to be used in targeting women in general, and women of influence and politically active women in particular.

- The study used graph theory to understand the Twitter network structure for each sample, to detect and measure distinct communities, and to understand the roles of users generating, amplifying and participating in the offensive speech. Offensive speech and the counter-narratives were annotated manually.

- The Twitter platform has been chosen for the study because the women have an active presence on it.

Findings

- Each woman has been mentioned hundreds of times with at least one of the terms in the lexicon (table 1). The offending accounts use the terms as a reaction to the targeted women's political opinions and/or alignment, and as a commentary on political opinions they expressed or were attributed to them on Twitter or outside the Twitter platform, such as in television appearances or print media. Specific examples are not being shown in this report, so that the hate content is not amplified.

- The main triggers of offensive content include contentions around Yemen's politics (Tawakkul Karman); Qatar's political rift with Bahrain, Egypt, Saudi Arabia and the UAE (Ghada Oueiss); politics in Lebanon (Dima Sadek); and women's rights activism (Loujain Hathloul).

---

14 The small-scale study is part of an ongoing project by the GBV Lab. The Lab uses a combination of social science, data science and graph theory to explore how the various elements of the digital ecosystem intertwine with gender inequality within development and humanitarian contexts.

	LOUJAIN HATHLOUL	TAWAKKOL KARMAN	DIMA SADEK	GHADA OUEISS
First Tweet	2013-01-23 T01:19:33.000 Z	2011-09-19 T08:47:36.000 Z	2014-03-16 T16:37:57.000 Z	2012-12-05 T18:40:37.000 Z
Last Tweet	2021-03-12 T16:01:33.000 Z	2021-12-03 T10:48:56.000 Z	2021-12-03 T19:21:16.000 Z	2021-11-12 T12:00:06.000 Z
	COUNT	COUNT	COUNT	COUNT
All tweets	472	3035	482	866
Different authors	450	2459	447	770
Retweets	241	1046	201	506
Replies	33	1250	128	180
Original Tweets	198	635	119	151
Quotes	0	104	34	29

**Table of Twitter data for the Sample of Arab women of influence on Twitter**

In addition to the offensive texts, some users posted sexually or racially offensive photos, demeaning the targeted woman and portraying her in a politically or socially offensive manner. **The photos are too offensive to include in this report, and only few blurred examples (to protect the dignity of the women) are included, to demonstrate the level of offensiveness.**

For example, a user posted a photo of the targeted woman describing her as a "whore" and invited users to like the post if they agreed with the description. The photo received more than 1,500 likes and was retweeted 101 times (exhibit 1). Other users posted photos of what appears to be the same woman bathing privately in a swimming suit (exhibit 2). Other users posted a photo of what they claimed to be one of the targeted woman in a bikini, yet wearing the headscarf, in public (exhibit 3). Some accounts use the subjects' names as hashtags to disseminate pornographic imagery.

Exhibit 1: The user posts a photo of Al Jazeera's Ghada Oueiss with the text: "If you think the label Al-Jazeera's Whore fits her, hit the like button". The post received 1,527 likes and was retweeted 101 times at the time of data retrieval.
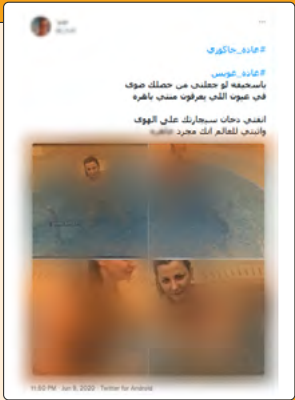
Exhibit 2: An account posting what appears to be private photos of Ghada Oweiss, using her name as a hashtag and describing her as a whore.

Exhibit 3: A post describing Tawakkol Karman as a prostitute and other vulgar descriptions, with a photo alleging to show her in a bikini and a headscarf.

- Nodes of influence emerged in each graph. Each node, or a Twitter account, is a creator of original offensive tweets or graphics that were retweeted, replied to or quoted favourably by other users (Graphs from 1 to 4). As a result, a distinct community is formed around each piece of content. Each community is a subset of nodes, densely connected to each other and loosely connected to the nodes in the other communities in the same graph. Each community is centred around a political affiliation opposing the targeted woman's political orientation.

Graphs from 1 to 4: Distinct communities around the offensive content that emerged in each woman's Twitter network.

1. DimaSadek Graph

2. GhadaOueiss Graph

**3. TawakkolKarman Graph**

4. LoujainHathloul Graph

- Some users post the offensive content and mention the targeted woman's Twitter username, apparently to draw her attention to the offensive content. Others post offensive content mentioning the targeted woman's name only. Others are more aggressive and use the woman's name and the offensive content as a hashtag, probably as an attempt to raise visibility and create a trend.

- Some of the offensive accounts use what appear to be real names, while others use pseudonyms.

- Counter-narratives, in the form of distinct communities opposing the use of offensive content, emerged in each woman's graph.[15] The counter-speakers condemned the use of sexual slurs against the woman, and some blamed this on the patriarchal system (exhibit 4). Others introduced hashtags in support of the targeted women. Some of the offensive accounts, however, hijacked the defending hashtags to post more offensive content and imagery (exhibit 5).

Exhibit 4: A post critical of the use of sexual slurs to target Dima Sadek. It reads: "If we have to be busy talking about someone like Dima Sadek and attack her, there is an endless list of issues. But because she is a woman and because our society is nasty and backward compared to other people, we immediately use the patriarchal system and filth. It has become normal that if our rival is a man then he is an opponent, but if it is a woman, she is certainly a whore."

---

[15] The counter-narrative emerged in the data because the counter-speakers used the query terms or replied to tweets with the offensive content in their defence.
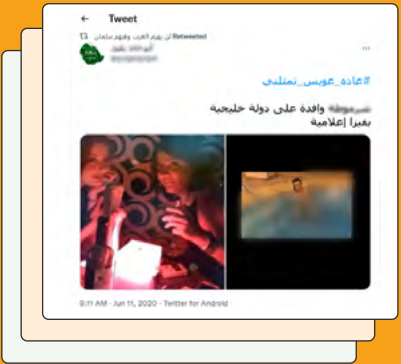
Exhibit 5: An account using a hashtag defending Ghada Oweish saying "Ghada Oweish represents me", in answer to a post with what appear to be private photos of her bathing in a swimming suit, and with the comment: "An expat prostitute in a Gulf country [Qatar] on a journalist visa."

Discussion

This report, albeit limited in scope, confirms the malicious relationship between digital space and violence against women. Some users choose to direct sexualised insults and sexist slurs at women over contentious political issues, instead of engaging in civil debates. It evidently takes not only a few original perpetrators, but a cohort to amplify the gender-based violence. It is disturbing to notice how perpetrators are determined to hurt and aggressively seek to raise the visibility of the offensive content by promoting the vile and obscene content as hashtags, and how they also mention the targeted woman's username to draw her attention.

While the textual and graphic hate content is unsettling and shocking, it speaks to the larger truth about GBV. Publishing them is not only a validation that can further bolster the narrative on GBV but is also a violation of the women's rights to protection and dignity. This research prioritises a do-no-harm approach.

Moving forward and recommendations

More large-scope studies are needed to explore the extent of hate speech in general, and GBV in particular in the Arabic Twittersphere and on other **social media platforms. The studies need to be** supplemented with contextually specific research that examines to what extent this type of hate speech causes women in the region to self-censor or withdraw completely from the platforms.

There is also a need to investigate whether violent speech discredits women, damages their reputations, and causes them emotional stress.

This type of research should mobilise CSOs to combat GBV, to educate women and girls about their rights to report GBV to the platforms, as well as about the related tools available to them, and to tackle the stigma associated with reporting sexual slurs directed at them. CSOs should also actively engage Twitter and other social media plateforms in reviewing and acting on GBV, and should hold them responsible for their inaction. For its part, Twitter should invest more in detecting hate content in the Global South languages and contexts and address its West-centric bias.

Cross-regional Recommendations

This report aims to present evidence-based recommendations inspired by the findings and contexts of each partner.

To enhance information integrity so citizens can form their opinions and vote without being exposed to manipulation campaigns, and prevent hate speech from being translated into violence, the report recommends that the relevant actors:

1. Agree on a definition of hate speech that does not curb the freedom of expression online:

One of the main challenges our partners faced during their social media monitoring was agreeing on one definition of hate speech that takes into consideration their local contexts, while respecting international standards. The main resources we found useful to our partners were the UN's Rabat Plan of Action and International Covenant on Civil and Political Rights.

2. Build networks and coalitions

The project launched its Words Matter network [16] to work with our partners in Sudan, Tunisia, Lebanon, and Jordan. Such networking efforts will facilitate the regional exchange of knowledge and information on the main behaviours of disinformation and hate speech in their countries.

3. Gather data and encourage open data practices, whereby data is collected while respecting personal data protection

- Another main challenge we faced is access to data. Social media platforms should give access to social media monitoring tools, such as CrowdTangle, to more researchers in MENA. Social media platforms should not only depend on western institutes and researchers to look into the Arabic content in MENA but should empower local researchers and invest in building their capacities

---

[16] The Words Matter network is also in contact with different information integrity bodies in the region and works closely with other human rights and content moderation coalitions in the region. The main aim of the coalition is to advocate for inclusive content-moderation policies and transparency practices when it comes to harmful content in Arabic.

to monitor social media in their local contexts.

- Researchers and regional CSOs should consider data ethics as the core of their practice. They should collect only qualitative public data while respecting personal data protection laws in their countries.

- Social media platforms should share with researchers from the region their data about hate speech and disinformation campaigns and their resolution, and information on how their content moderation system works in Arabic.

- Researchers and CSOs should build user-friendly tools, methodologies and resources to detect hate speech and different information manipulation in Arabic across different dialects and contexts, in collaboration with social media platforms and funders.

4. Better coordinate with social media platforms on content removals

According to our partners' experience, the detection of harmful content is challenging when social media platforms take some items down and remove content. There should be better coordination between civil society groups monitoring social media and tech companies to make clear on what basis the content was taken down, which is one of the principles of transparency and accountability in content moderation laid down by The Santa Clara Principles.[17]

---

[17] More information on The Santa Clara principles can be found at: https://santaclaraprinciples.org/

About Words Matter

DRI has been increasingly active in the field of social media monitoring (SMM) since 2017, strengthening local capacities to monitor social media during elections, sharing information and evidence gathered in different countries, bringing together expert organisations, producing methodologies, and informing public and expert debate.

Within the framework of the project "Words Matter", DRI and its partners seek to contribute to strengthening the safeguarding of democratic processes and societies' resilience to online disinformation and hate speech in the MENA region.

DRI works with partner organisations from four countries (Jordan, Lebanon, Sudan and Tunisia), strengthening local capacities to monitor and analyse online disinformation and hate speech during key national democratic processes, while building a regional network to allow for comparative analysis and peer learning.

"Words Matter" aims to achieve the following objectives:

– Capacity-building for project partners **to acquire institutional skills to design sound social media monitoring methodologies**, to effectively monitor disinformation and hate speech online, and to enhance evidence of the impacts of disinformation and hate speech online on civic or political participation and human rights.

– **Enhanced multi-stakeholder and regional engagement** to advocate against and combat online disinformation and hate speech, through a civil society network, as well as through continuous exchanges on transparent regulations.; and

– In the countries of project partners, **improved awareness and resilience of civic target groups**, and concrete action by decision-makers to transparently combat online hate speech and disinformation.

About Digital Democracy

DRI's Digital Democracy (DD) programme protects online democratic discourse by exposing information, manipulation and hate speech, strengthening the capacity of CSOs for monitoring and advocacy, and ensuring appropriate and evidence-based responses from governments and tech companies.

DRI is well-positioned to address online threats and disinformation, due to its research on manipulated media content, deepfakes as potential disinformation tools, and its current focus on identifying new potential threats and emerging technologies in this field. As part of our diverse toolbox, we have, for example, integrated machine learning models to help us identify emerging trends in the disinformation space. Our work on information manipulation is also complemented by analysing and publishing guides on gender-based under-representation and harassment online.

An important activity of the DD programme in exposing and fighting hate speech and disinformation is social media monitoring (SMM). SMM is the objective analysis of democratic discourse and political actors on social media platforms. Social media monitoring is far more complex than traditional media monitoring, with a myriad of actors and content, combining official democratic institutions (e.g., parties, politicians, media) and unofficial actors (e.g., individuals, political influencers, partisan groups). This is why DRI published the Digital Democracy Monitor Toolkit, the first social media monitoring methodology that helps civil society, journalists and academia to research social media and democracy.

Our methodology was tested and used for conducting social media monitoring in 12 countries (including Germany, Libya, Myanmar, Nigeria and Sri Lanka), focusing on disinformation, hate speech and political advertising before, during and after the elections. By using a holistic approach to analyse social media, our toolkit engages with disinformation and hate speech by looking at the message or content, the active messengers, and the messaging, thus both the forms and the channels of distribution.

Based on the findings of our SMM, we have advocated for the implementation of the European Democracy Action Plan (EDAP) commitments, which could strengthen the fight against disinformation at the EU level, and contributed to the debate about content ranking systems, a major challenge when it comes to the dissemination of dis/misinformation. DRI has also lobbied for the implementation of the EU's Digital Service Act, a potential milestone in the effort to increase accountability across social media platforms. In launching the Arabic version of the SMM toolkit, we hope to empower the MENA region in the same way.

**DEMOCRACY
REPORTING
INTERNATIONAL**

**Democracy Reporting International
(DRI) was founded in 2006 by an
international group of experts
on democratic governance and
elections.**

**DRI works on research and
analysis to direct engagement
with partners on the ground to
improve democratic structures and
safeguards across the countries
where we work.**