

# « Without Interference »: State Sponsored Digital Disinformation and the Corporate Responsibility to Respect

Candidate number: 8011

Submission deadline: 16 May 2022

Number of words: 20064



## Acknowledgements

Foremost, I would like to express my sincere gratitude to the Fritt Ord Foundation, which awarded this project the Fritt Ord Student Stipend on human rights, the right to free speech, and/or democracy (Spring 2022).

I would also like to give my warmest thanks to my supervisor, Professor Bård Anders Andreassen, whose guidance and feedback carried me through all the stages of writing this project.

I would also like to acknowledge the members of the Human Rights and Sustainable Development: Law and Policy Research Group at the Norwegian Centre for Human Rights for their advice and insightful feedback.

I wish to extend my special thanks to my interviewees who readily provided invaluable insight and perspectives on a very much sensitive subject, and without whose participation this project would have not been possible.

My sincere thanks also go to my fellow Master candidates for their support, commiseration and camaraderie throughout this last semester.

Last but certainly not least, I would like to thank my mother for her unwavering support and sympathetic ear.

# Table of contents

- 1 INTRODUCTION..... 1**
- 1.1 Background ..... 1
- 1.2 Objective and Scope..... 5
- 1.3 Relevance and Justification..... 6
- 1.4 Research Questions ..... 7
- 1.5 Methodology and Sources..... 8
- 1.6 Structure ..... 10
- 2 FACEBOOK AND THE CORPORATE RESPONSIBILITY TO RESPECT  
FREE SPEECH..... 11**
- 2.1 The ‘respect’ pillar applied to Facebook’s activities ..... 11
  - 2.1.1 A gap in the conversation ..... 11
  - 2.1.2 A content-based impact ..... 13
- 2.2 The specificities of free speech on social media..... 15
  - 2.2.1 The dilemma of content regulation..... 15
  - 2.2.2 Users as business relationships..... 19
- 2.3 The Facebook Responsibility Scale ..... 21
- 3 THE MADAGASCAR CASE ..... 24**
- 3.1 Background and context..... 24
- 3.2 Case mapping and analysis ..... 26
  - 3.2.1 Mapping process..... 26
  - 3.2.2 Campaigns presentation..... 30
  - 3.2.3 General Findings..... 36
- 3.3 Impacts analysis and Facebook’s involvement..... 41
- 4 FACEBOOK’S RESPONSES: IN MADAGASCAR, AND BEYOND..... 45**
- 4.1 An array of untapped measures..... 45
  - 4.1.1 Case-specific measures, or lack thereof ..... 45
  - 4.1.2 Corporate Human Rights Policy..... 46
  - 4.1.3 Due Diligence on Disinformation..... 48
  - 4.1.4 Voluntary Measures on Disinformation ..... 52
- 4.2 An insufficient implementation ..... 53
  - 4.2.1 Deflecting Responsibility ..... 54
  - 4.2.2 Understating Leverage Capacity ..... 55
  - 4.2.3 North-South Disparities ..... 57

4.3	Recommended measures .....	59
4.3.1	UNGP measures .....	59
4.3.2	Beyond UNGP? .....	60
<b>5</b>	<b>CONCLUSION.....</b>	<b>62</b>

**List of tables**

Table 2.1. Facebook responsibility scale.....23

Table 3.1. Individually Targeted Disinformation Cases in Madagascar.....40

Table 3.2. Facebook responsibility scale – Degree of involvement.....44

Table 4.1. Overview of independent HRIAs commissioned by Facebook.....50

**List of figures**

Figure 3.1. The Anatomy of state-sponsored trolling model.....29

## List of abbreviations

AI	Artificial Intelligence
BHR	Business and Human Rights
BSR	Business for Social Responsibility
CVO	Covid Organics
EU	European Union
GNI	Global Network Initiative
HRD	Human Rights Defenders
HRDD	Human Rights Due Diligence
HRIA	Human Rights Impacts Assessment
ICCPR	International Covenant on Civil and Political Rights
ICT	Information and Communication Technology
ILO	International Labor Organization
OECD	Organisation for Economic Co-operation and Development
OHCHR	Office of the United Nations High Commissioner for Human Rights
RFI	Radio France International
RSF	Reporters Sans Frontières
TIM	Tiako I Madagasikara
TVM	Televiziona Nasionaly Malagasy
UN	United Nations
UNGP or GP	United Nations Guiding Principles or Guiding Principles
WSJ	Wall Street Journal

# 1 Introduction

## 1.1 Background

In October 2021, France24 published a report accusing authorities in Madagascar of bankrolling fake profiles on Facebook.com to share biased or false information for political gain. The report features anonymous testimonies from alleged state-sponsored “trolls” describing the processes and goals behind these practices. For example, one woman, tasked with creating fake profiles and commenting on publications posted by a team leader, showed France24 journalists a fake publication where the cook of a popular restaurant poses with the Malagasy president, to convey an image of closeness to the people. According to her, these operations are led by the Minister of Communication, who denied any wrongdoing.<sup>1</sup> Besides praising the government, “trolls” would also criticize and discredit journalists and opposition leaders.<sup>2</sup>

The advent of the internet and new information and communication technologies (ICT) has brought significant changes in our communication processes and given rise to new societal issues<sup>3</sup>. The current digital landscape harbors growing levels of disruptive and manipulative communications, especially over social media platforms.<sup>4</sup> Much of the literature and debate on the subject tends to generalize the “widespread exploitability of information on the internet”<sup>5</sup> under the epithet of “Fake News” but the situation referenced above, and the deliberate emphasis on “criticize and discredit” over other goals, mirrors a specific trend of states such as Ecuador,<sup>6</sup> the Philippines,<sup>7</sup> and Venezuela<sup>8</sup> using tactics of digital disinformation as part of broader attacks against perceived dissenting voices.<sup>9</sup>

Often, such campaigns target individual journalists, activists, and others who criticize the government or its affiliates.<sup>10</sup> One notable example was the campaign launched against Maria Ressa, founder of the Filipino media outlet Rappler and co-recipient of the 2021 Nobel Peace

---

<sup>1</sup> France 24, “Facebook 'troll farms'.”

<sup>2</sup> *Id.*

<sup>3</sup> Bennett & Livingston, “A Brief History,” 3. Bradshaw & Howard “The Global Organization,” 1.

<sup>4</sup> Bennett & Livingston, *id.*

<sup>5</sup> Helm & Nasu, “Regulatory Responses”, 302.

<sup>6</sup> Woolley, Monaco & Nyst, “State-sponsored trolling”, 29.

<sup>7</sup> *Id.*, 32.

<sup>8</sup> *Id.*, 41.

<sup>9</sup> *Id.*, 5.

<sup>10</sup> Woolley, Monaco & Nyst, “State-sponsored trolling,” 12.



Prize<sup>11</sup>, after she reported on her government’s misuse of social media in 2016.<sup>12</sup> As in the Madagascar story, the attacks against Ressa occurred predominantly on Facebook.<sup>13</sup> Affirming the company’s passive awareness, Ressa denounced Facebook, Inc.<sup>14</sup> as either “negligent” or “complicit in state-sponsored hate.”<sup>15</sup>

Ressa’s accusation is a thorny one. It sparks an intuitive acquiescence as social media companies are being increasingly put under criticism for their response to what is viewed as harmful user content, often being accused of platforming hate speech and spreading disinformation. However the appropriate response to harmful media content is complicated by the existence of countervailing rights of freedom of expression.<sup>16</sup>

Article 19 of the International Covenant on Civil and Political Rights (ICCPR) guarantees the right to hold opinions without interference<sup>17</sup> and to seek, receive and impart information and ideas of all kinds, regardless of frontiers and through any media.<sup>18</sup> The United Nations Human Rights Council and General Assembly stated that the freedoms people enjoy offline also apply online.<sup>19</sup> While freedom of opinion (the right to hold opinions) is absolute, freedom of expression (the right to seek and impart information and ideas) may be restricted under certain circumstances.<sup>20</sup> But under international human rights law the falsity of information alone cannot be a legitimate ground for restrictions,<sup>21</sup> as freedom of expression extends to the right to impart information and ideas that “may shock, offend and disturb.”<sup>22</sup> However, the 2017 Joint Declaration on Freedom of Expression and “Fake News”, Disinformation and Propaganda contends that “State actors should not make, sponsor, encourage or further disseminate statements which they know or reasonably should know to be false (disinformation).”

---

<sup>11</sup> The Nobel Prize, “The Nobel Peace Prize.”

<sup>12</sup> Woolley, Monaco & Nyst, “State-sponsored trolling”, 32.

<sup>13</sup> *Id.*, 33.

<sup>14</sup> Facebook, Inc. has changed its corporate name to “Meta” in October 2021 as part of a large rebrand campaign. For the purposes of this paper, we will refer to the company by its former name. BBC, “Facebook changes.”

<sup>15</sup> Woolley, Monaco & Nyst, “State-sponsored trolling”, 32.

<sup>16</sup> Joseph, “The Human Rights Responsibilities,” 6.

<sup>17</sup> ICCPR, art. 19.

<sup>18</sup> *Id.*

<sup>19</sup> GA 68/167; UNHRC A/HRC/RES/26/13; UNHRC A/HRC/ RES/32/13; UNHRC A/HRC/32/38.

<sup>20</sup> ICCPR, art. 19(3).

<sup>21</sup> GC CCPR/C/GC/34, para 49.

<sup>22</sup> “Joint Declaration,” para 7.

Under the ICCPR, States have a duty to refrain from interfering with free speech, and commandeering online information to intimidate and silence critical individuals constitutes interference.<sup>23</sup> In her 2021 report on ‘Disinformation and freedom of opinion and expression’, the Special Rapporteur Khan distinguishes “state-sponsored disinformation” as a key concern which “can emanate from State institutions directly or from proxies targeting audiences”<sup>24</sup> and which can aim to “[...] curb protests against and criticisms of Governments”<sup>25</sup>. Bradshaw and Howard refer to “cyber troops” as “state-sponsored organizations tasked with conducting disinformation campaigns on the Internet”<sup>26</sup>. Nyst and Monaco dub the phenomenon as “state-sponsored trolling”<sup>27</sup> Wu notes a diminished reliance on direct censorship as a means of speech control,<sup>28</sup> in favor of new techniques such as “trolling” which “seek to humiliate, harass, discourage, and even destroy targeted speakers using personal threats, embarrassment, and ruining of their reputations”<sup>29</sup>.

Whether instigated by States or non-State actors, ideologically or financially motivated, online disinformation campaigns often involve social media companies, as platforms have become a key medium for many people in the world to receive and impart information. While disinformation can be traced as far back as the invention of the printing press,<sup>30</sup> the scale to which digital technology has enabled it to be created and disseminated is unprecedented and poses serious challenges to democracy and human rights.<sup>31</sup> According to Benesch, Facebook by itself is running more human communication than any government does, or ever has.<sup>32</sup>

Benesch underlines that social media companies hold “extraordinary, transnational power and influence”<sup>33</sup> and “human rights impacts of great magnitude”<sup>34</sup> and that yet they “govern largely on their own, and in secret.”<sup>35</sup> This reflects the core issue of business and human rights,

---

<sup>23</sup> Woolley, Monaco & Nyst, “State-sponsored trolling”, 46.

<sup>24</sup> *UNHRC A/HRC/47/25*, para 47.

<sup>25</sup> *Id.*, para 49.

<sup>26</sup> Bradshaw & Howard “The Global Organization,” 1.

<sup>27</sup> Woolley, Monaco & Nyst, “State-sponsored trolling,” 1.

<sup>28</sup> Wu, “First Amendment,” 558.

<sup>29</sup> *Id.*, 560.

<sup>30</sup> Politico, “The Long and Brutal History.”

<sup>31</sup> *UNHRC A/HRC/47/25*, para. 2.

<sup>32</sup> Benesch, “But Facebook’s,” 86.

<sup>33</sup> *Id.*, 92.

<sup>34</sup> *Ibid.*

<sup>35</sup> *Ibid.*

which Biltchiz dubs “the problem of corporate human rights obligations.”<sup>36</sup> The quest for filling the “accountability gap” of business-related human rights abuses has led to many regulatory initiatives, both in soft and hard law, most predominant of which being the “Protect, Respect, Remedy” framework of the U.N. Guiding Principles on Business and Human Rights (UNGP).

The UNGPs provide the principal international instrument defining the basic norms of responsible corporate behavior. They assign to corporations a responsibility to respect human rights,<sup>37</sup> primarily those rights enshrined in the International Bill of Rights and the International Labor Organization’s (ILO) Declaration on Fundamental Principles and Rights at Work, including article 19 ICCPR. The UNGPs indicate that companies ought to avoid infringing on those rights and address adverse human rights impacts with which they are involved.<sup>38</sup> A crucial means of fulfilling these responsibilities is the conduct of “human rights due diligence” (HRDD).<sup>39</sup>

The responsibility to respect constitutes a social expectation rather than a legal duty, the UNGP themselves falling under the category of “soft” rather “hard law.” Under this framework, corporations lack any binding positive obligations in relation to human rights, as can be seen through the deliberate use of “responsibility” rather than “obligation.”<sup>40</sup> Deva deplores this approach as a “regressive step”<sup>41</sup> which “unduly narrowed down the scope of corporate obligations.”<sup>42</sup> In his view, companies should in certain situations have independent obligations such as the ‘protect’ and ‘fulfill’ types of duties along with states.<sup>43</sup> Bilchitz criticizes the UNGPs as inadequate and unduly restrictive, arguing that the normative foundations of key international human rights instruments do incur legally binding and wider obligations on corporations.<sup>44</sup> Recently, domestic measures introduced in various States address this issue, with countries such as France,<sup>45</sup> Germany<sup>46</sup> and Norway<sup>47</sup> adopting mandatory human rights

---

<sup>36</sup> Biltchiz, “Putting Flesh on the Bone,” 2.

<sup>37</sup> OHCHR HR/PUB/11/04, Principle 11 & commentary.

<sup>38</sup> *Id.*

<sup>39</sup> OHCHR HR/PUB/11/04, Principle 1.

<sup>40</sup> UNHRC A/HRC/14/27.

<sup>41</sup> Deva, *Regulating Corporate*, 109.

<sup>42</sup> *Id.*, 110.

<sup>43</sup> *Ibid.*

<sup>44</sup> Biltchiz, “A chasm between,” 108.

<sup>45</sup> Herbert Smith Freehills, “Respecting Human Rights.”

<sup>46</sup> Herbert Smith Freehills, “Hurdles on the Journey.”

due diligence laws, while a proposed EU-wide mandatory due diligence framework<sup>48</sup> is being negotiated. Another relevant initiative is the ongoing negotiation process for a “Legally Binding Instrument to Regulate, in International Human Rights Law, the Activities of Transnational Corporations and Other Business Enterprises”<sup>49</sup> at the UN level, though consensus has yet to be reached.

Social media companies play a major role in the critical challenges generated by disruptive user content such as disinformation, and their efforts to address the issue have been deemed inadequate and detrimental to human rights<sup>50</sup>. In her aforementioned report, Khan called for a recalibration of State responses to disinformation along with a review of the business models of digital platforms.<sup>51</sup>

## 1.2 Objective and Scope

This project aims to contribute to the academic debate on the scope and ambit of the corporate responsibility to respect human rights, focusing on the harm generated through social media platforms to the right to freedom of expression as enshrined in article 19 ICCPR. To a lesser extent, the research purports to assess how well suited the UNGPs are as a framework to guide corporate responses to challenges to the digital right to freedom of expression. To achieve this, the research will undertake a normative-empirical analysis on the nature and scope of Facebook, Inc.’s corporate responsibilities regarding the right to freedom of expression and the information content published on its Facebook platform.

The research will analyze the issue of state-sponsored digital disinformation in Madagascar (the Madagascar case) through a business and human rights (BHR) approach. Whereas some authors encompass the concepts of “disinformation” and “misinformation” under the term of “fake news”<sup>52</sup>, Nuñez underlines the importance of distinguishing between the three terms to address their unique specific threats.<sup>53</sup> For the purposes of this research, we will refer to the Wardle and Derakhshan classification of types of information disorders<sup>54</sup> to distinguish be-

---

<sup>47</sup> LOV-2021-06-18-99.

<sup>48</sup> Herbert Smith Freehills, “EU Mandatory Human Rights.”

<sup>49</sup> *UNHRC A/HRC/RES/26/9*, 2-3.

<sup>50</sup> *Id.*, para. 3.

<sup>51</sup> *Id.*, para. 5.

<sup>52</sup> Helm & Nasu, “Regulatory Responses”, 302.

<sup>53</sup> Nuñez, “Disinformation Legislation”, 785.

<sup>54</sup> Wardle & Derakhshan, “Information Disorder.”

tween misinformation (when false information is shared, but no harm is meant), disinformation (when false information is knowingly shared to cause harm) and malinformation (when genuine information is shared to cause harm). The research will focus on “disinformation” understood as the subset of misinformation that is knowingly propagated.

### 1.3 Relevance and Justification

Disinformation has been studied from the perspectives of the media, technology, or politics,<sup>55</sup> but viewing it through a BSR approach to freedom of expression enables unique insights and relevance.

Disinformation in the digital era is so pervasive that Livingston and Bennett shortcut to speaking of the “Disinformation Age” as governments spend significant resources to manipulate the public debate via social media.<sup>56</sup> In this regard, Wu notes the end of the “scarcity of information” era under which censorship was the main threat to free speech and the advancement in its stead of a new paradigm where information is plentiful and may be used to attack and silence as much as it is used to illuminate or debate.<sup>57</sup> Wu argues that in the 21st century, danger comes less from directly suppressive states and more from states that “targets listeners or undermines speakers indirectly” by weaponizing speech itself.<sup>58</sup>

Under this new paradigm of threats to free speech, social media companies often appear as a key actor. These companies face critical human rights dilemmas: combating what is viewed as harmful content risks silencing speech that, under international law, should be permitted.<sup>59</sup> Moreover, social media platforms do not constitute an empty page on which every internet user has an equal right to write, but rather they manipulate the dissemination of information according to commercial imperatives, prioritizing high engagement through often controversial material.<sup>60</sup>

Facebook’s surveillance-based model, in particular, has been subjected to scrutiny after whistleblower Frances Haugen leaked internal documents showing the company’s profit oriented

---

<sup>55</sup> Nunez studied disinformation as a distortion to the “*marketplace of ideas*”. Wu studied disinformation as weaponized speech.

<sup>56</sup> Bradshaw & Howard, “Troops, Trolls and Troublemakers”; see also Bradshaw & and Howard, “Challenging Truth and Trust”.

<sup>57</sup> Wu, “Is the First Amendment”, 558.

<sup>58</sup> *Ibid.*

<sup>59</sup> UNOHCHR, “Moderating online content.”

<sup>60</sup> Woolley, Monaco & Nyst, “State-sponsored trolling,” 48.

management of the real-world harm it causes.<sup>61</sup> The company's subsequent rebrand to Meta has been surmised as an effort to distance the brand from these revelations. In March 2021, eight months before the Facebook Papers, Facebook released a Corporate Human Rights Policy for the first time in its history. Although a necessary step, the initiative has been met with skepticism in the wake of the company's past and ongoing failures in safeguarding data, respecting free expression, and protecting vulnerable users. Moreover, the policy seems to raise more questions than answers about how Facebook will integrate human rights considerations into its day-to-day operations.

At the same time, the conversation around due diligence often centers around labor rights, environmental and climate change impact assessments, with few studies applying it to "regular human rights" and fewer even to a free speech perspective. An approaching endeavor would be Benesch's article proposing international human rights law as a framework for social media companies to regulate speech and be held accountable,<sup>62</sup> though the article does not offer a significant focus on the responsibility to respect. Thus this project offers the opportunity to interpret a specific human rights concept under a new context.<sup>63</sup>

#### **1.4 Research Questions**

This project undertakes to address Ressa's accusation: has Facebook been negligent, or is it complicit? This research will transpose the question to the Madagascar case and rephrase it as such: to what extent is Facebook, Inc. responsible for harms to the right to freedom of expression caused by state-led disinformation campaigns deployed on their Facebook platform in Madagascar?

The main question will be divided into three sub-questions: What are Facebook's responsibilities regarding freedom of expression and its information content? What measures has the company taken to address its involvement in these impacts? What additional measures, if any, should the company take in the context of the case? This research will endeavor to as accurately as possible situate Facebook on a responsibility scale going from not responsible to complicit, and derive valuable insights on the human rights responsibilities of corporations operating digital platforms.

---

<sup>61</sup> AP News, "People or Profit."

<sup>62</sup> Benesch, "But Facebook's," 91.

<sup>63</sup> Andreassen & Vihn, "Business' Duties," 1.

## 1.5 Methodology and Sources

The research is designed as a normative-empirical analysis. The research question will be answered within the context of the Madagascar situation taken as a case study. Although case studies are a well-used method in human rights research<sup>64</sup> as they offer a convenient “unit of observation,”<sup>65</sup> we acknowledge the limitations of single-country analysis for valid generalization and do not aim for a full-fledged comparative nor causal-explanatory study. Rather, we focus on a specific case to derive a point of departure for theoretical interpretation.<sup>66</sup>

The research will combine desk-based documentation with qualitative insights obtained through semi-directed interviews with key stakeholders such as civil society representatives and individuals who have been targeted by state-led disinformation for opinions they hold or ideas they imparted that could be perceived as a form of dissent.

First, the research will map a selection of six cases of individually targeted disinformation campaigns spread on the Facebook platform from 2018 to 2021 in Madagascar. Each case reflects an emblematic aspect of the issue, for example Case 1 concerning a smear campaign against a journalist reporting on a famine in the South of Madagascar while Case 2 considers attacks against a civil society member. The timeframe parallels president Rajoelina’s current term<sup>67</sup> to anchor the case in a more cohesive context. The mapping aggregates specific disinformation content into concerted and coherent disinformation campaigns through social media scraping. The campaigns will be analyzed against a model “anatomy of a state-trolling campaign” developed by Nyst and Monaco<sup>68</sup> to isolate such criteria as target identity, triggers, and language used.

A key point in understanding disinformation is measuring its impact, and therefore assessing which goals it pursued and whether said goals were achieved.<sup>69</sup> This requires access to the minds of disinformation producers who in practice seek to conceal their identity<sup>70</sup>. The most methodologically rigorous studies of disinformation compensate this gap through years of

---

<sup>64</sup> Andreassen, “Comparative analysis,” 246.

<sup>65</sup> *Id.*, 239.

<sup>66</sup> *Id.*, 247.

<sup>67</sup> BBC, “Madagascar ex-leader.”

<sup>68</sup> Greene et al, “Mapping Fake News,” 20-21

<sup>69</sup> *Id.*

<sup>70</sup> *Ibid.*

massive data collection.<sup>71</sup> Given the scope of this project, such an approach would not be feasible. Instead we rely on perception data collected through interviews and triangulated with social media scraping<sup>72</sup>. Therefore, we have to acknowledge the limitations of perception data throughout the research implementation.

Following the principle of non-maleficence,<sup>73</sup> research participants will be interviewed on the basis that their responses would not be attributed to them and that their affiliations would not be identified. Interviewees give voluntary and informed consent to participate in the research<sup>74</sup> through a consent form describing the goals of the research and what consent entails, such as video recording and withdrawal possibility.<sup>75</sup> Following the principle of confidentiality,<sup>76</sup> respect for privacy in data collection will also be given the utmost consideration.

Second, the research will map Facebook's responses with regard to harmful disinformation content published on its platform. These responses may have preceded the selected campaigns - as part of company policies - or succeed them - as part of either acknowledgment or denial of responsibility. Here, the mapping process shall rely on a combination of primary sources such as documentation on relevant content regulation rules as well as secondary sources such as press releases, news entries, and academic articles.

Finally, the research will involve a document analysis of discourse<sup>77</sup> around the nature and scope of the corporate responsibility to respect human rights and how it applies to the digital right to freedom of expression and to the Madagascar case. For this purpose, the analysis will rely on the text of international human rights law, including hard and soft law instruments, notably the UNGPs and the ICCPR. The research shall also benefit from reference to secondary sources, in the form of NGO and institutional reports, critical doctrinal contributions and other types of scholarly work<sup>78</sup>.

---

<sup>71</sup> *Ibid.*

<sup>72</sup> Bowen, "Document Analysis," 28.

<sup>73</sup> Ulrich, "Research Ethics," 196.

<sup>74</sup> *Id.*, 199.

<sup>75</sup> Nygaard, *Writing your master's*, 39.

<sup>76</sup> Ulrich, "Research Ethics," 199.

<sup>77</sup> *Id.*

<sup>78</sup> Bowen, "Document Analysis," 32.



## **1.6 Structure**

Chapter 2 will set out the theoretical framework surrounding the subject. It will reflect the corporate responsibility to respect human rights (Pillar 2) and how and to what extent it applies to social media businesses such as Facebook, given the inherent dilemmas of free speech regulation. The chapter will propose a UNGP-based “responsibility scale” for instances of adverse human rights impacts of Facebook related to the right to freedom of expression.

Chapter 3 will provide an empirical analysis of the Madagascar case. The chapter will present six selected cases of allegedly state-sponsored disinformation campaigns launched on the Facebook platform between 2018 and 2021 against individuals sharing information or ideas critical of the current Malagasy regime. Drawing from Chapter 2, we will then determine whether and to what degree Facebook has been involved in adverse impacts to free speech.

Chapter 4 will then map and assess the relevant Facebook responses in the context of the case to provide a normative analysis of the soundness and efficacy of these measures. To answer the main research question, the chapter will then explore the desirability of an extended regime of corporate human rights responsibility for social media companies.

## **2 Facebook and the Corporate Responsibility to Respect Free Speech**

This chapter discusses Facebook's corporate responsibilities regarding the right to freedom of expression, to provide the theoretical foundations to guide further reflection. More precisely, this chapter aims to determine the contours of the corporate responsibility to respect human rights regarding free speech infringements involving user content on the Facebook platform.

### **2.1 The 'respect' pillar applied to Facebook's activities**

#### **2.1.1 A gap in the conversation**

The UNGP mark a milestone in the quest for accountability for human rights abuses.<sup>79</sup> The human rights regime developed in the twentieth century crystallized states as sole duty bearers, yet by the 1990s multinational corporations had reached a level of societal influence that transcended the reach of public governance. This "accountability gap" provided opportunities for companies to commit harmful acts without sanction or redress and spurred a quest for regulation first through binding instruments such as the United Nations Code of Conduct on Transnational Corporations in 1974 or the UN Draft Norms on the Responsibilities of Transnational Corporations and Other Business Enterprises in 2003.<sup>80</sup> Other developments consist in soft law, such as the OECD Guidelines for Multinational Enterprises in 1976, the ILO Tripartite Declaration of Principles Concerning Multinational Enterprises in 1977, the UN Global Compact in 2000, and the UNGP in 2011.

The UNGP adapt existing human right standards for market actors<sup>81</sup> and provide the first internationally accepted standard for preventing and addressing adverse human rights impacts linked to business activity. They encompass three pillars outlining how states and businesses should act:

- The state duty to protect human rights against harm caused by third parties;
- The corporate responsibility to respect human right;

---

<sup>79</sup> Pillay, "The corporate responsibility."

<sup>80</sup> Though it was met with stiff resistance and garnered little international support.

<sup>81</sup> Andreassen & Vinh, "Business' Duties Across Borders," 7.

- Actions that both states and companies must perform to enhance access to remedy for victims.

The UNGP have been said to put more emphasis to human rights' ethical and moral character than their potential legal ramifications.<sup>82</sup> They provide that the duty of states under international law to protect human rights against third party-caused violations must be complemented by corporate commitment to respect human rights in practice, regardless of the effectiveness of states' actions. Since their implementation, the GPs have generated a wide array of national and international measures, issued by governments as well as business associations. Human rights groups, workers organizations and civil society are also using the GPs as an advocacy tool. Despite criticisms regarding their normative reach, the GPs are recognized as state of the art in terms of articulating corporate responsibilities for human rights.<sup>83</sup>

The GPs emphasize equal application to all business enterprises regardless of size, sector, location, ownership, or structure. However, Joseph notes that certain industries, such as extractive companies, the garment industry, and pharmaceutical corporations, have been subjected to greater scrutiny.<sup>84</sup> The focus on these industries is understandable given their capacity for serious human rights impacts, but the UNGP's key premise remains that all industries can generate impacts, both positive and negative, on the enjoyment of all human rights.<sup>85</sup>

One sector which has been overlooked according to Joseph is the media industry, a situation she deems a "glaring gap in the conversation" given the societal impact of platforms such as Facebook.<sup>86</sup> Social networking sites have revolutionized human communication<sup>87</sup> and Facebook by itself centralizes more than 2 billion active users worldwide.<sup>88</sup> One illustration of Facebook's potential to negatively impact human rights would be its role in the ethnic cleansing of Myanmar's Muslim Rohingya population.<sup>89</sup> In the ICT sector, the Global Network Initiative Principles (GNI Principles) do offer principles on privacy and freedom of expression

---

<sup>82</sup> O'Brien & Dhanarajan, "The Corporate Responsibility," 2.

<sup>83</sup> Wettstein, "Human Rights as a Critique," 18.

<sup>84</sup> Joseph, "The Human Rights Responsibilities," 1.

<sup>85</sup> *Id.*

<sup>86</sup> *Ibid.*

<sup>87</sup> UNHRC CCPR/ C/GC/34, para. 15. See also O'Flaherty, "Freedom of Expression."

<sup>88</sup> Statista, "Number of monthly active".

<sup>89</sup> See BSR, "Human Rights Impact."

with guidelines referring to the UNGPs.<sup>90</sup> However, the GNI Principles focus on protecting those rights against governmental interference rather than protecting individuals against human rights infringements facilitated by social media companies.<sup>91</sup>

### 2.1.2 A content-based impact

The GPs set the baseline responsibility of all enterprises to respect human rights wherever they operate, independently of the ability or willingness of states to fulfill their own obligations, and over and above compliance with national laws and regulations.<sup>92</sup> Businesses must comply with national laws with respect to human rights,<sup>93</sup> but they should also self-regulate to ensure that they respect human rights in the absence of adequate domestic regulation or enforcement.<sup>94</sup> Another foundational principle states that businesses should seek to prevent or mitigate adverse human rights impacts that they have “caused or contributed to”, as well as those “directly linked” to their operations, products or services through their business relationships, whether contractual or non-contractual.<sup>95</sup>

The UNGPs set out the steps for companies to identify, prevent, mitigate, and remedy their adverse impacts through the process of due diligence.<sup>96</sup> This process differs from the financial or legal liability due diligence, as it identifies impacts to people rather than to corporate assets.<sup>97</sup> Beyond legal compliance, the UNGPs focus on the need to prevent and address adverse human rights impacts<sup>98</sup> by embedding human rights commitments and policies into the company’s day-to-day operations.

The due diligence process concerns human rights impacts that a company causes<sup>99</sup>, those to which the company contributes and those to which the company’s operations, products, or services are directly linked by its business relationship with another entity.<sup>100</sup> In such situa-

---

<sup>90</sup> GNI, “GNI Principles,” para. 2.

<sup>91</sup> Buhmann & Olivera, “Human rights and social media,” 8.

<sup>92</sup> *UNHRC A/HRC/17/31*, para. 55; *OHCHR HR/PUB/11/04*, Principle 11 & Commentary.

<sup>93</sup> *UNHRC A/HRC/17/31*, *id.*

<sup>94</sup> *Id.*, para. 58; *OHCHR HR/PUB/11/04*, Principle 12.

<sup>95</sup> *Id.*, Principle 13.

<sup>96</sup> *OHCHR HR/PUB/11/04*, Principles 17–22.

<sup>97</sup> *Id.*

<sup>98</sup> Ruggie & Sherman, “The Concept of ‘Due Diligence’,” 923.

<sup>99</sup> *OHCHR HR/PUB/11/04*, Principle 17(a).

<sup>100</sup> *Id.*, incl. commentary.

tions, a company that is able to should exercise “leverage” to help prevent or mitigate its impacts.<sup>101</sup> In the BHR context, leverage refers to the influence a business enterprise has with an entity involved in a human rights abuse, for example a government or a company in the value chain.<sup>102</sup> In practice, leverage may include a broad range of measures, which enterprises may undertake by themselves or together with others.<sup>103</sup>

The UNGPs’ due diligence approach has been integrated into regulatory instruments such as the EU Corporate Sustainability Due Diligence Directive and into national legal instruments such as the French Corporate Duty of Vigilance Law, the Norwegian Transparency Act and the German Supply Chain Act. The 2011 revision of the OECD Guidelines<sup>104</sup> adopted human rights due diligence and, using the term risk-based due diligence, expanded its application beyond human rights to labor relations, environmental issues, anti-corruption, and consumer concerns. The OECD issued several sector-specific guidance documents, as well as a guidance document for risk-based due diligence that applies to all sectors, though no specific guidance has been developed for the ICT sector yet. For application and interpretation purposes, these OECD texts complement the UN’s BHR instruments<sup>105</sup> along with the OHCHR’s general guidance on the corporate responsibility to respect human rights in 2012<sup>106</sup>.

As any other business, Facebook has a responsibility to avoid causing or contributing to harm to human rights in relation, for example, with its workforce or environmental impacts. However, Joseph notes the particularity of social media’s core business, which is the publication and dissemination of content.<sup>107</sup> The media industry in general might employ workers and impact the environment, but the product/service which constitutes a media business’s main output and income sets them apart from other sectors. Joseph further distinguishes between social media and mainstream media in that the former do not publish their own content, rather they publish and curate content created and posted by others.<sup>108</sup> Gillespie defines social media

---

<sup>101</sup> *Id*, Principle 19.

<sup>102</sup> *Id*, Principle 19, incl. commentary.

<sup>103</sup> *Id*, Principles 17–19, incl. commentaries.

<sup>104</sup> OECD, “OECD Guidelines.”

<sup>105</sup> Bağlayan, “Searching for Human Rights,” 371. Buhmann, “Analysing OECD,” 391.

<sup>106</sup> OHCHR, “The Corporate Responsibility.”

<sup>107</sup> Joseph, “The Human Rights Responsibilities,” 1.

<sup>108</sup> *Id*, 10.

platforms as sites and services which “host, organize, and circulate” user content.<sup>109</sup> Lévesque further argues that “curation is the product” rather than simply content.<sup>110</sup> The business model of social media sites will be further discussed in 2.2, but this content-grounded model implies content-grounded potential human rights impacts.

Furthermore, the UNGP refer to the concept of “saliency” to guide enterprises in identifying potential adverse impacts. The most salient human rights for a business enterprise are those that stand out as being most at risk and in practice will be the ones a company prioritizes.<sup>111</sup> Saliency will vary according to the sector and operating context. As a company operating on user content, Facebook’s corporate responsibilities appear by way of logic most likely to arise from said user content. We will thus focus on the areas where Facebook, as a company, is pursuing its “core business”: the provision of a platform where users can consume, publish and/or disseminate content. We also limit our discussion to user content possibly infringing on freedom of expression. In contrast, we will not be focusing on those areas where Facebook as a company is acting beyond the scope of the operations which distinguish it from other non-social media businesses, such as abuses regarding its workforce.

## **2.2 The specificities of free speech on social media**

### **2.2.1 The dilemma of content regulation**

Applying the corporate responsibility to respect to the right to freedom of expression in relation to Facebook’s user content poses the problem of content moderation.

On the one hand, article 19(2) ICCPR provides the right to freedom of expression for everyone “regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of his choice” and the jurisprudence of the Human Rights Committee applies the provision to freedom of expression on the internet.<sup>112</sup> Means of expression are considered to include “all forms of electronic and internet-based modes of expression”.<sup>113</sup> In the terms of the ICCPR, freedom of expression can only be restricted for protection of “the rights or reputations of others”, “national security or of public order (*ordre public*), or of public health or

---

<sup>109</sup> Gillepsie, *Custodians of the Internet*.

<sup>110</sup> Lévesque, “Applying the UN Guiding,” 7.

<sup>111</sup> *Id.*, 8.

<sup>112</sup> See *UNHRC A/HRC/17/27* and *UNHRC A/HRC/RES/20/8*, para 1.

<sup>113</sup> GC 34, para. 12.

morals” and only given that the restrictions are “provided by law” and “necessary”.<sup>114</sup> Any restrictions on the operation of websites, blogs or any other internet-based, electronic or other such information dissemination system, including systems to support such communication, such as internet service providers or search engines, are only permissible to the extent that they are compatible with Article 19(3).<sup>115</sup> Compulsory restrictions on expression are imposed by Article 20, which mandates the prohibition of propaganda for war and hate speech on racial, religious or national grounds. As it stands, Facebook content can fall under protected speech under article 19(2) ICCPR as well as in the zone of speech that should be punished for compromising the rights of others.

On the other hand, the UNGP state that businesses, including social media companies, have an independent responsibility to respect human rights, which stems from “a global standard of expected conduct for all business enterprises wherever they operate”.<sup>116</sup> Corporations should avoid causing or contributing to adverse human rights impacts through their activities and address such impacts when they occur. For the Facebook operations which interest this paper, this means avoiding causing or contributing to those adverse impacts resulting from content posted on the Facebook platform. Facebook user content can threaten privacy, security of the person, a person’s freedom from hate speech, and other rights. As social media users as a group lack traditions or expectations of ethical behavior,<sup>117</sup> social media posts are likely to ignore injunctions imposed for privacy reasons,<sup>118</sup> to “blurt out” prejudicial details about an accused person in an upcoming trial,<sup>119</sup> or to constitute hate speech.<sup>120</sup>

So what is the appropriate response to harms to human rights arising from or facilitated by social media content? The answer is complicated by the existence of countervailing rights of freedom of expression, as measures taken by social media providers to prevent one type of

---

<sup>114</sup> ICCPR, art. 19(3).

<sup>115</sup> *UNHRC CCPR/C/GC/34*, para. 43.

<sup>116</sup> *UNHRC A/HRC/17/31*, 13.

<sup>117</sup> Joseph, “The Human Rights Responsibilities,” 10.

<sup>118</sup> For example, in 2012, a superinjunction prevented the naming of a footballer allegedly having an affair. Social media named the footballer, and eventually the superinjunction was lifted as it was no longer viable. See BBC, “Ryan Giggs .”

<sup>119</sup> See Bell, “Social media,” remarking on the sharing on social media of prejudicial information of a man facing trial in Melbourne for murder.

<sup>120</sup> Meixler, “UN fact-finders.”

harm (for example, the offence of child pornography) can themselves lead to harmful human rights impacts. Joseph distinguishes two issues regarding social media companies.<sup>121</sup>

The first issue concerns the extent to which a company complies with domestic censorship laws. Facebook, for example, has long been negotiating to enter the Chinese market, and has signaled a willingness to comply with China's censorship demands.<sup>122</sup> Any human rights abuses in this instance are driven by the State, rather than the social media company. One could argue that in such a situation Facebook would be complicit in State human rights abuse. However, it is arguable that the social media company is not doing any harm, as the restrictions on free speech would exist regardless of their involvement. Pillar 2 does not direct businesses to disobey domestic laws even if those laws might compel human rights abuses.

The second issue concerns company-imposed censorship, which gives rise to Pillar 2 considerations. There are few official remedies available if Facebook, for example, chooses to take one's page down.<sup>123</sup> The reconstruction of a page on another platform, particularly if the page had thousands of followers and sophisticated multimedia, is difficult if not impossible.<sup>124</sup> In this regard, the Fake News Declaration states that social media companies: "[...] should respect minimum due process guarantees including by notifying users promptly when content which they created, uploaded or host may be subject to a content action and giving the user an opportunity to contest that action, subject only to legal or reasonable practical constraints, by scrutinising claims under such policies carefully before taking action and by applying measures consistently."<sup>125</sup>

Social media posts could also allude to complicity. "Complicity in the BHR context refers to the indirect involvement of companies in human rights abuses"<sup>126</sup> and has legal and non-legal meanings.<sup>127</sup> Joseph argues that media content which promote abusive policies can "be seen as" contributing to a government's decision to adopt or retain, and therefore implement, that

---

<sup>121</sup> Joseph, "The Human Rights Responsibilities."

<sup>122</sup> The Guardian, "Facebook may 'block content'".

<sup>123</sup> See Morozov, "The Net Delusion," 213.

<sup>124</sup> In the leadup to the Arab Spring, a page called "We are all Khaled Said" was taken down due to the use of a pseudonym by the administrator (for safety reasons). The page had thousands of subscribers and much multimedia content. It was eventually restored. See Joseph, "The Human Rights Responsibilities," 177.

<sup>125</sup> "Joint Declaration," 4(c).

<sup>126</sup> *UNHRC A/HRC/8/16*, para. 30.

<sup>127</sup> *OHCHR HR/PUB/11/04*, Principle 17 incl commentary.



policy.<sup>128</sup> However, for social media content, the main issue lies in the appropriate liabilities of social media businesses regarding content posted by users.

States take differing approaches to this question: the US excludes intermediary liability<sup>129</sup> while the EU provides some but not total immunity to the business running the platform.<sup>130</sup> The European Court of Human Rights discussed this issue in *Delfi AS v Estonia* and ruled that intermediary liability for third party posts is not per se a breach of the right to freedom of expression.<sup>131</sup> In contrast, the Fake News Declaration has stated that: “[i]ntermediaries should never be liable for any third party content relating to those services unless they specifically intervene in that content or refuse to obey an order adopted in accordance with due process guarantees by an independent, impartial, authoritative oversight body ... to remove it and they have the technical capacity to do that”.<sup>132</sup> Here, reliance is put on Pillar 1 to solve the issue – censorship is to be enforced by State bodies rather than the platforms themselves. However, the Declaration also recognizes that self-regulation may have a role to play: “Where intermediaries intend to take action to restrict third party content (such as deletion or moderation) which goes beyond legal requirements, they should adopt clear, pre-determined policies governing those actions. Those policies should be based on objectively justifiable criteria rather than ideological or political goals and should, where possible, be adopted after consultation with their users”.<sup>133</sup>

Online platforms have revolutionized the social conditions of speech.<sup>134</sup> Combining lower costs for generating and sharing information with access to a wide array of interaction, the largest platforms such as Facebook and YouTube exert considerable influence over public discourse.<sup>135</sup> Though these platforms have tended to deny the extent of this influence,<sup>136</sup> it is undeniable they “govern speech”<sup>137</sup> through an array of content curation methods designed to

---

<sup>128</sup> Joseph, “The Human Rights responsibilities,” 5.

<sup>129</sup> See Communications Decency Act 1996, s 230 (US).

<sup>130</sup> See EU E-Commerce Directive 2000/31/EC.

<sup>131</sup> See *Delfi AS v Estonia*. See also Spano, “Intermediary Liability,” 665.

<sup>132</sup> “Joint Declaration,” principle 1d.

<sup>133</sup> “Joint Declaration,” principle 4a.

<sup>134</sup> Balkin, “Digital Speech,” 2.

<sup>135</sup> Sander, “Freedom of Expression,” 941.

<sup>136</sup> See Ash et al, “GLASNOST!”.

<sup>137</sup> Benesch, “But Facebook’s,” 86.

both “gatekeep” and “organize” what their users see and hear.<sup>138</sup> Gillespie’s definition of social media platforms includes that they “moderate the content and activity of users”.<sup>139</sup>

Thus platforms – an ambiguous term oftentimes reclaimed by companies to inspire neutrality and evade regulatory scrutiny<sup>140</sup> - at their core do far more than platform: they also edit, often for profit and growth, to the extent that it has spurred concerns about corporate imperatives overriding public interest or silencing legitimate debate.<sup>141</sup> In 2018, David Kaye proposed to re-align the private incentives of platform governance with larger public interests through “a framework for the moderation of user-generated online content that puts human rights at the very center.”<sup>142</sup> But even a human rights-based approach to online content moderation cannot evade the regulatory pitfalls inherent to the right to freedom of expression.<sup>143</sup>

### 2.2.2 Users as business relationships

One way to activate Pillar 2’s regime of responsibility that has been proposed by authors is to interpret social media users, who provide the bulk of the content making up the commodity of the social media platforms, as business relationships in the sense of the UNGPs.

Companies need revenue to make profit and remain in the market. A business model expresses how a company works in terms of its value proposition (the desirability of the goods or services it offers), value creation and provision, and financial endurance (costs and revenue).<sup>144</sup> The conventional business model for companies providing goods to individuals as consumers is based on the consumer obtaining the company’s product for a monetary price. A clothing company, for example, obtains revenue by producing or sourcing a piece of clothing and selling it to consumers against cash or cash equivalents.<sup>145</sup>

The business model of social media providers that give access to platforms without charging a fee fundamentally differs from the conventional model. This model makes it seem like users have free access to a product – the platform – but in fact themselves, or rather the data they

---

<sup>138</sup> Sander, “Freedom of Expression,” 941.

<sup>139</sup> Gillespie, *Custodians of the Internet*.

<sup>140</sup> Caplan, “Content or Context Moderation.”

<sup>141</sup> Jørgensen, “Human Rights and Private Actors,” 251-53.

<sup>142</sup> *UNHRC A/HRC/38/25*, para. 2.

<sup>143</sup> See Sander “Freedom of Expression.”

<sup>144</sup> Rohn, “Social Media Business Models.”

<sup>145</sup> Buhmann & Olivera, “Human rights and social media,” 11.

represent, are the revenue. Social media users “pay” for access to platforms with data about their buying patterns, activities, and preferences. Social media providers track, store and analyze that data, using it for targeted advertising and selling it to other companies. This provides the social media provider with its revenue.<sup>146</sup> For these companies, “user data is the currency and curation the product”.<sup>147</sup>

Scholars distinguish between two types of social media companies. The first comprises social media providers that own and operate social media platforms. Facebook is one such company. The second comprises companies using platforms for their communication with customers and other stakeholders. The latter pay companies of the first category for access to user data.<sup>148</sup> This can be seen to bring the user into a “business relationship” with the social media provider as the user, by supplying data, creates a source of revenue for the provider.

In other words, when social media companies provide access to platforms, they provide a service to which users supply data in return. Under this paradigm, users transcend the status of consumers, or stakeholders, to attain that of business relations, within the meaning of the UNGPs.<sup>149</sup> This adds a new dimension to the potential of companies to contribute to harmful impacts. It also has implications for HRDD and the appropriate actions companies must take when they are involved in a human rights infringement.<sup>150</sup>

Differentiating between causing a risk or harm, contributing to one, or being directly linked to one can be difficult, and the connections involving social media providers and their stakeholders are often complex.<sup>151</sup> For example, a Facebook user can share photos that infringe upon the privacy of others. Buhman and Olivera compare this to an enterprise using cobalt obtained through child labor in its products or an investor providing capital to an investee company and thereby enabling it to function.<sup>152</sup> As the harmful impact is directly linked to the social media company through the service provided and the user’s actions, the company should, at the least, exercise leverage. That leverage needs to be appropriate to the context and

---

<sup>146</sup> *Id.*

<sup>147</sup> Lévesque, “Applying the UN Guiding Principles,” 7.

<sup>148</sup> Rohn, “Social Media Business Models.”

<sup>149</sup> Lévesque, “Applying the UN Guiding Principles,” 7.

<sup>150</sup> Buhmann & Olivera, “Human rights and social media,” 12.

<sup>151</sup> *Id.*, 12.

<sup>152</sup> *Id.*, 13

the target group: in such a case, the actor causing harm is an individual acting in a non-profit capacity. For these purposes, leverage might entail targeted efforts by the social media provider, such as building users' capacity for identifying situations that may result in human rights risks.<sup>153</sup>

Alternatively, the social media provider could be contributing to the risk or harmful impact by providing the technical solution through which the infringement occurs. The user's action causes the impact, but the company enables it. This mirrors the situation of a company lending vehicles to security forces that use the vehicles to travel to local villages and commit atrocities, where in case the 'vehicle' provided is the platform itself. From an impact by omission perspective, it could also be said that the user causes the harm, but by failing to act on its knowledge of users' harmful use of its platform and the potential of its platform to enable such use, the social media provider can be seen to have contributed to that harm.<sup>154</sup>

### **2.3 The Facebook Responsibility Scale**

While analyzing tensions between the responsibility to respect free speech of social (and mainstream) media businesses and the countervailing rights to freedom of expression, Joseph concludes that "there is no easy answer"<sup>155</sup> as to the appropriate response expected from Facebook under the Ruggie process for free speech infringements involving user content. As many other issues in the human rights field, the issues surrounding Facebook's free speech responsibilities can be interpreted as less of a dichotomy – to be resolved by a binary of choices: to platform or to edit – and more of a tension – to be managed through a balance of circumstances and imperatives. As such, this paper proposes the following 'responsibility scale' as a reflective tool to proceed with the paper's research goals and to contribute to general scholarship on the issue.

The proposed 'Facebook responsibility scale' is adapted from the OHCHR and OECD guidance texts on forms of business involvement in harmful impacts and appropriate actions. Based on the UNGPs, these texts identify that companies can be involved with human rights impacts by:

---

<sup>153</sup> *Ibid.*

<sup>154</sup> Buhmann and Olivera, "Human rights and social media," 13.

<sup>155</sup> Joseph, "The Human Rights Responsibilities," 7.

- **Causing them:** in this case, the company’s activities are sufficient on their own to result in the adverse impact and it would be required to take appropriate action to stop these activities, to mitigate and remedy any ongoing harm, and prevent or mitigate future harm.<sup>156</sup>
- **Contributing to them:** in this case, the company’s actions cause, facilitate or incentivize another entity to cause an adverse impact. This is mainly relevant for substantial contributions, not those that are minor or trivial<sup>157</sup> and requires for the company to cease, remedy, prevent or mitigate, and/or use its leverage to influence the entity causing the harm. The UNGPs however recognize that situations requiring leverage can be complex, in terms of what would be appropriate action for the company to take to meet its responsibility to respect human rights.<sup>158</sup>
- **Being directly linked through operations, products, or services by its business relationship with another entity:** this category proves complex, both in terms of assessing the character of the linkage and determining what would constitute appropriate action.<sup>159</sup> Direct linkage is defined by the relationship between the harm and the company’s operations, products, or services through a ‘business relationship’.<sup>160</sup> An additional layer of complexity arises in our case as mentioned above as to whether social media users can be analyzed as a business relationship to the social media provider in the sense of the UNGPs. For the enterprise that is directly linked to the harm, appropriate action entails using its leverage to prevent or mitigate the harm.<sup>161</sup>

For the purposes of this research, the scale extends over and above the UNGP’s classification to include levels of non-involvement and grounds for complicity, to better delineate the varying possible situations in which Facebook user content may or may not present negative impacts on free speech. The scale is presented in this chapter will then be applied to our case analysis in Chapter 3.

---

<sup>156</sup> *OHCHR* HR/PUB/11/04, 41; OECD, “OECD Due Diligence,” 46–47.

<sup>157</sup> *Id.*, 46

<sup>158</sup> *OHCHR* HR/PUB/11/04, Principle 19, incl. commentary.

<sup>159</sup> *OHCHR* HR/PUB/11/04, Principle 19, incl. Commentary.

<sup>160</sup> OECD, “OECD Due Diligence,” 46–47.

<sup>161</sup> *Id.*, 27; OECD, “Supply Chains,” 14.

**Table 2.1: Facebook Responsibility Scale**

Adapted from the UNGP's three ways companies can be involved with human rights impacts—cause, contribution and linkage—which define how companies are expected to respond. Designed for cases of free speech infringements involving the Facebook platform's user content.

Level of involvement	Non-impacting	Non-involved	Caused	Contributed	Linked	Complicit - non-legal	Complicit - legal
<b>General definition</b>	No human rights impact	Human rights impacts registered in the periphery of the company's activities but not engaging the company's responsibility under Pillar 2	Causes an impact through its own activities	Contributes to an impact either directly or through some outside entity (government, business or other)	A company's operations, products or services are linked to a negative human rights impact through a business relationship (or series of relationships)	Indirect involvement in human rights abuses	Aiding and abetting to human rights impacts National criminal liability or International criminal law jurisprudence: "knowingly providing practical assistance or encouragement that has a substantial effect on the commission of a crime".
<b>Facebook-specific definition</b>	Facebook's activities involving its user content do not infringe on the right to freedom of expression of others	Facebook's activities involving its user content restricts freedom of expression in conformity with international human rights law standards (Article 19(3) and 20 ICCPR)	Facebook directly infringes on freedom of expression with regard to user content	Facebook contributes to free speech impacts directly or through an other entity	Facebook is directly linked to a free speech infringement through the service its provide and its users' action	Facebook indirectly and knowingly causes/enables abuses to free speech by a third party	Facebook is aiding and abetting to unlawful restrictions on freedom of expression
<b>Scenario</b>	Facebook platforms posts in a way and content that does not infringe on freedom of expression	Facebook applies content moderation policies with respect to the demands of freedom of expression, and with satisfactory oversight and transparency	Facebook applies content moderation policy which generally targets and suppresses posts on certain subjects or by certain groups, in opposition with international standards for freedom of expression	Facebook contributes to harm to free speech by enabling harmful posts (providing the technical solution or failing to acknowledge and prevent harmful use of its platform)	Facebook users coordinate harmful posts on the company's platform without Facebook leveraging against it	Facebook allows/facilitates its platform to be used in harmful ways by a third party	Facebook allows/facilitates its platform to be used in unlawful and harmful ways by a third party
<b>Appropriate actions</b>	None	None	<ul style="list-style-type: none"> <li>• Cease the activity that caused the impact</li> <li>• Provide remedy</li> <li>• Take steps to prevent impact from recurring</li> </ul>	<ul style="list-style-type: none"> <li>• Cease activity and avoid contribution</li> <li>• Provide remedy</li> <li>• Use leverage to mitigate any remaining impact to the greatest extent possible</li> </ul>	Has forward-looking responsibility to prevent the impact from recurring <ul style="list-style-type: none"> <li>• No explicit responsibility to provide remedy but can choose to do so</li> </ul>	-	-
<b>Conditions</b>	-	-	-	Users as business relationship	Users as business relationship	Facebook is knowingly involved Facebook's contribution is not minor or trivial	Facebook is knowingly involved Facebook's contribution is not minor or trivial Legal liability is provided
<b>Example</b>	-	Hate speech or war propaganda is disseminated on the Facebook platform and removed in conformity with relevant international and domestic norms, and with sufficient transparency as part of Facebook's due diligence process.	Napalm Girl case: Facebook removing a post by a Norwegian journalist containing an iconic photograph of a girl fleeing a napalm attack during the Vietnam War.	Facebook's algorithm prioritises maximized user engagement over other considerations and facilitates the spread of disinformation campaigns	Myanmar Case	Facebook establishes an informal cooperation with state agencies which encourage the company to respond to removal requests within narrow timeframes, effectively enabling the State to circumvent legal avenues to remove alleged illegal content.	

### 3 The Madagascar case

This chapter analyzes instances of individually targeted, state-sponsored disinformation campaigns launched on Facebook in Madagascar. The empirical foundation thus provided, along with the UNGP-based theoretical findings of Chapter 2, will enlighten our assessment of Facebook’s level of involvement the resulting adverse free speech impacts.

#### 3.1 Background and context

Madagascar’s Constitution provides for the right to freedom of expression and freedom of the press.<sup>162</sup> The country has also ratified the ICCPR and its optional protocols. Still, the country scores relatively low on well-established reports such as those of Freedom House<sup>163</sup> or Reporter Sans Frontières (RSF)<sup>164</sup>. Precisely, freedom of expression in Madagascar is challenged through a two-fold mechanism, both in law and in practice.

First, free speech guarantees in Madagascar are undermined by criminal libel laws and other restrictions. The 2016 Code of Media Communication<sup>165</sup> decriminalizes press offenses but swapped criminal penalties for disproportionately high fines.<sup>166</sup> Journalists and civil society members have decried the law as “liberticide” and strategically dissuasive.<sup>167</sup> The law conceded welcomed modifications in 2020,<sup>168</sup> though how these legal changes transfer into practice remains to be seen. The 2016 law on cybercriminality<sup>169</sup> prohibits online “defamation” or “insult” toward a public or institutional figure and practice has shown authorities to adopt a wide interpretation of these offenses.<sup>170</sup>

---

<sup>162</sup> Madagascar Constitution of 2010, art. 10.

<sup>163</sup> Madagascar scored 60/100 in 2021, qualifying as “Partly free”. Freedom House, “Freedom in the World.”

<sup>164</sup> Madagascar scored 57 in RSF’s the 2021 World Press Freedom Index. RSF, “Madagascar.”

<sup>165</sup> Communication Code, Loi n° 2016-029.

<sup>166</sup> Africa News, “New Controversial Code.”

<sup>167</sup> Midi Madagasikara, “Code de la Communication.”

<sup>168</sup> Communication Code, Loi n°2020-006.

<sup>169</sup> Cybercriminlaty Code, Loi n°2016-031.

<sup>170</sup> For example, the case of investigative journalist Fernand Cello arrested in 2017 and acquitted by the Fianarantsoa Court of Appeal nearly two years later for reasons related to his professional activities and following fabricated charges of check book theft. Amnesty International, “Un journaliste relaxé.” See also: Jeune Afrique, “Rasoanaivo” and RSF, “RSF dénonce.”

Second, this legal stronghold on expression and information is exacerbated by the particularities of Madagascar's media landscape.<sup>171</sup> Studies mapping media ownership in Madagascar underline the high percentage of media outlet owned by politicians or businessmen with a prominent political obedience.<sup>172</sup> This results in “a concentrated and unbalanced media landscape,”<sup>173</sup> where journalists are at best challenged and at worst constrained by their editorial lines.<sup>174</sup> This situation climaxed during the 2020 covid-19 pandemic, when the government took drastic measures to control information shared by the media and individuals. It evoked Law 91-011 of 1991 – which was intended to be applied in exceptional circumstances – and permitted the media to share only official government information. Other restrictive measures included prohibiting radio stations from broadcasting phone-in show.<sup>175</sup> The government targeted and harassed journalists and media outlets who allegedly spread “false information” about pandemic responses.<sup>176</sup>

According to our interviewees, these aspects concur to foster an environment of “self-censorship”, where journalists and citizens feel constrained in how they can express and impart information.<sup>177</sup> Social media platforms' potential for speech liberation and democratization, for example by overriding remoteness and bringing forth voices and topics that would otherwise have been “reserved to bars and bistros”,<sup>178</sup> thus offer an opportunity to leverage the situation for the better. This potential has been demonstrated after the passage of a destructive cyclone in 2022, as instances of “online citizen journalism” have contributed to disprove misleading official reports on the extent of destruction and the efficacy of authorities' responses.<sup>179</sup> This aspect of platforms' societal impact reinforces the relevance of questioning the free speech responsibilities of Facebook, especially as in Madagascar disparities between

---

<sup>171</sup> As well for freedom of press as safety risks involved in the investigation of sensitive subjects such as cattle rustling and the illicit extraction and sale of natural resources. RSF, “Madagascar.”

<sup>172</sup> FES Madagascar, « Les impacts de la structure. »

<sup>173</sup> Politika, “Etude sur le monde des medias.”

<sup>174</sup> Eg: a journalist working independently on a story about the contents of the covid-19 remedy developed and promoted by state authorities in 2020 has received a warning from their chief editor that they will bear the consequences alone if their work comes to reflect badly on the media owner to the presidency. Communication to author, 5 May 2020.

<sup>175</sup> Amnesty International, “Madagascar.”

<sup>176</sup> Freedom House, “Freedom in the World.”

<sup>177</sup> Interview 3.

<sup>178</sup> Interview 2.

<sup>179</sup> *Id.*



expensive internet data bundles and cheap Facebook packages means that for a lot of people, “Facebook is the internet”.<sup>180</sup>

## 3.2 Case mapping and analysis

### 3.2.1 Mapping process

This section aims to establish the empirical basis for assessing Facebook’s corporate responsibility. The scope of this research does not permit for an exhaustive account of online disinformation in Madagascar is not feasible. Rather, we picked a selection of cases to showcase the phenomenon through a process of inductive reasoning.<sup>181</sup> These cases were identified through social media scraping triangulated with what interviewees deemed as salient cases. We acknowledge the limitations of this approach, such as the reliance on perception data and the fact that the inductive approach can only infer probable (rather than true) conclusions.<sup>182</sup> For the purposes of this research, state responsibility is therefore understood, much as in corporate involvement as per the UNGP, as a spectrum of degrees of involvement, rather than an evidence-based causality. This approach might be less reliable but it is a well-established proxy, as attributing responsibility for actions that occur in the online realm is at best imperfect and at worst impossible.<sup>183</sup> According to the Atlantic Council, “the Internet enables anonymity more than security,” thus policy makers often struggle to determine the source of cyber-attacks while analysts often fall into the trap of “attribution fixation”.<sup>184</sup>

There have been various studies on online disinformation campaigns, state-sponsored or not,<sup>185</sup> such as Bradshaw and Howard’s 2018 paper<sup>186</sup>, itself drawing from their 2017 Computational Propaganda Project.<sup>187</sup> However, Nyst and Monaco’s typology appears most relevant as it applies specifically to state-led social media campaigns designed to repress dissent. In their 2018 report, Nyst and Monaco conducted case studies of “state-sponsored trolling” across countries such as the United States, Azerbaijan, Bahrain, Ecuador, the Philippines,

---

<sup>180</sup> *Id.*

<sup>181</sup> Walch, *Assessment Strategies for Science: Grades 6–8*, 4.

<sup>182</sup> Copi, Cohen, Flage, *Essentials of Logic*, 8.

<sup>183</sup> Woolley, Monaco & Nyst, “State-sponsored trolling”, 17.

<sup>184</sup> Healy, “Beyond Attribution,” 1-2.

<sup>185</sup> It is be noted that though these studies have been conducted across several countries, there is little literature or data available about Madagascar.

<sup>186</sup> Bradshaw & Howard, “The Global Organization.”

<sup>187</sup> Bradshaw & Howard, “Troops, Trolls and Troublemakers.”

Turkey, and Venezuela. The report shows that though the phenomenon occurs in a variety of contexts, it often deploys similar strategies.

In this paper, we will use the model developed by Nyst and Monaco to identify “state-sponsored trolling” and apply it to Facebook posts published between 2019 and 2021 in Madagascar to determine whether and to what extent the Madagascar experience compares to abuses observed for example in the Philippines of Duterte. The “in Madagascar” criterion does not refer to a strict geographical delimitation, as campaigns might target people living abroad as well, but refers to Facebook content that was posted by a user residing in Madagascar or references events happening in Madagascar.

The model surmises that state-sponsored attacks rise beyond general exploitation of digital infrastructures as a distinct category of disinformation campaigns. Others have used the term “patriotic trolling” to refer to these campaigns which often obscure, by design, the state’s role therein.<sup>188</sup> The term mirrors those used to describe state hacking campaigns carried out under the guise of independent hackers to mask the attacks’ provenance.<sup>189</sup> The common features of a state-sponsored digital campaign, according to Nyst and Monaco, are:

- **Targets and triggering actions:** journalists, activists, and others who criticize the government, government affiliates, or status quo institutions are the prime targets of states using digital platforms and tools.<sup>190</sup>
- **Language:** although state-trolling represents one of the new tactics of state control with “speech itself as a censorial weapon”<sup>191</sup>, online campaigns fall back on well-established messaging tactics to seed distrust and turn public opinion against journalists and activists<sup>192</sup> such as accusations of collusion with foreign intelligence agencies, accusations of treason, use of violent hate speech, creation of elaborate cartoons and memes, and politicians and their proxies using claims of “fake news” as a form of dog whistling to state-sponsored trolls.<sup>193</sup>

---

<sup>188</sup> ABS-CBN News, “Rappler Boss”; Geybulla, “In the Crosshairs.”

<sup>189</sup> Deibert & Rohozinski, “Liberation vs Control.”

<sup>190</sup> Woolley, Monaco & Nyst, “State-sponsored trolling,” 12.

<sup>191</sup> Wu, “Is the First Amendment,” 561-62.

<sup>192</sup> Woolley, Monaco & Nyst, “State-sponsored trolling,” 12.

<sup>193</sup> *Id.*, 13.

- **Use of bots and algorithms:** campaigns appropriate and game the algorithms of social media sites<sup>194</sup> to increase the prominence and pervasiveness of their messaging.<sup>21</sup>
- **Election antecedents:** the infrastructure and mechanisms for state sponsored trolling attacks in numerous countries have grown out of, or been built upon, infrastructure and mechanisms established during election campaigns. Once a candidate or party is successful, these same resources are often deployed in pursuit of consolidating and extending power.<sup>195</sup>

This typology is completed by Nyst and Monaco's "mechanisms of state responsibility", which differentiate the degree to which attacks can be attributed to the state. Nyst and Monaco agree with the Atlantic Council that in analyzing online disinformation, the question of who did it should be trumped by the question of who is to blame, and therefore designed a model resting on inference of probability rather than certainty of state responsibility. The model proposes a spectrum of overlapping state responsibility mechanisms:

- **State executed:** state-funded and -directed "cyber militia" execute strategies designed by the government to disseminate propaganda, isolate dissenting views, and drown out or remove anti-government sentiment.<sup>196</sup> These campaigns can be deployed through volunteers, amateurs and professionals.<sup>197</sup>
- **State directed or coordinated:** governments directing or coordinating, but not executing, online attacks, using coordination channels to outsource campaigns to private actors.<sup>198</sup>
- **State incited or fueled:** governments maintain a distance from the attack but both instigates and profits from it. Governments may use high-profile proxies and other government stand-ins to signal state support for a particular attack.<sup>199</sup>

---

<sup>194</sup> Gaming of algorithms is the deliberate exploitation of a platform's underlying automated process to achieve an outcome not intended by the platform.

<sup>195</sup> Woolley, Monaco & Nyst, "State-sponsored trolling," 13.

<sup>196</sup> Woolley, Monaco & Nyst, "State-sponsored trolling," 18.

<sup>197</sup> Henochowicz, "Youth Volunteers."

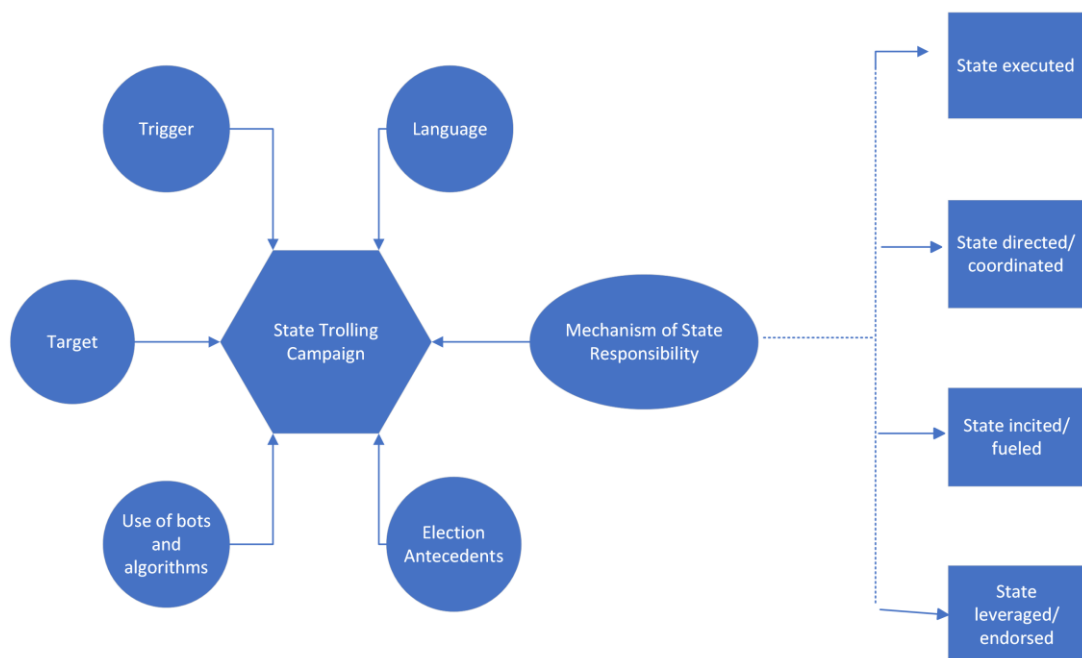
<sup>198</sup> Woolley, Monaco & Nyst, "State-sponsored trolling," 19

<sup>199</sup> *Id.*, 19.

- **State leveraged or endorsed:** governments point to the existence of seemingly independent groundswells of public opinion to justify and legitimate state positions. In doing so, it signals to internet users its tacit approval of harassment campaigns and implicitly promises impunity for state-sponsored trolls<sup>200</sup>

Through social media scraping, we aggregated Facebook posts and analyzed them against this model. For the purposes of this research, we only perform this form of discourse analysis on initial posts. Though we acknowledge the impact of comments and shares in building the “severity” of online attacks<sup>201</sup>, those will be left out of our discussion they often mirror initial posts in terms of content.

Figure 3.1: The Anatomy of State-Sponsored Trolling Model



Source: Adapted from Woolley, Samuel C., Nicholas Monaco and Nyst, Carly. “State-sponsored trolling: How Governments Are Deploying Disinformation as Part of Broader Digital Harassment Campaigns”. Report. Institute for the Future, 2018, pp 11-21. Diagram by author.

<sup>200</sup> *Id.*, 20

<sup>201</sup> Joseph, “The Human Rights Responsibilities,” 11; *UNHRC A/HRC/47/25*, para. 2.

## 3.2.2 Campaigns presentation

### 3.2.2.1 Case 1: Journalist

On 21 June 2021, independent journalist and international correspondent Gaelle Borgia posted a video<sup>202</sup> on her Facebook account showing locals in Ambovombe-Androy (South of Madagascar) eating scraps of leather from zebu skin, normally used to make sandals. The sequence illustrates the precariousness of the victims of a famine which affects part of the south of Madagascar.<sup>203</sup>

The video spread on social media and provoked chain reactions to question the probity of the reporter. First, in an official statement on 23 June 2021, the governor of the Androy region accused the journalist of “peddling false information” and exploiting the misfortune of the population “to try to shine publicly.”<sup>204</sup> Then, a report on the national TV channel (TVM), which is linked in its operation to the Ministry of Culture and Communication, accused the journalist of hiring the locals to make false testimonies.<sup>205</sup> In a new video shot and broadcast by Borgia, the witnesses in questions later admitted to having been threatened by men armed with a knife and bought to confirm these allegations. According to RSF, the TVM report was produced to discredit the journalist, who was not questioned to defend her version of the facts.<sup>206</sup> Both these tactics can be interpreted, following Nyst and Monaco’s model, as a way for the state to instigate and dog-whistle<sup>207</sup> online trolls through its proxies (the national TV channel and the governor in office).

Following the initial video publication by the journalist, and in parallel to the discredit campaign launched by the governor of the Androy region and the TVM channel, online attacks spread on the Facebook platform. A joint declaration from the foreign press correspondents in Madagascar noted that the journalist was the target of “a smear campaign, repeated insults on

---

<sup>202</sup> Facebook, “21 June 2021.”

<sup>203</sup> Amnesty International, “Global Leaders” and “Urgent Humanitarian Intervention.” See also Un News, “World Cannot Look Away.”

<sup>204</sup> RSF, “Une journaliste victime.”

<sup>205</sup> Facebook, “29 June 2021.”

<sup>206</sup> RSF, “Une journaliste victime.”

<sup>207</sup> A subtly aimed political message intended for, and can only be understood by, a particular demographic group.

social networks, which amount to cyberbullying.”<sup>208</sup> It is to be noted that posts attacking the credibility and professionalism of the journalist have been happening prior to the triggering event, however not to the same intensity. Facebook posts accused the journalist of orchestrating a fake news campaign to destabilize the current regime and disparage the image of Madagascar.<sup>209</sup> She is referred to as “the one who denigrates the regime with lies and misinformation”<sup>210</sup> and is accused of harboring a malicious “hatred” against the president.<sup>211</sup> Other posts accuse her of working to destabilize Madagascar for the benefit of foreign donors<sup>212</sup> and other states<sup>213</sup>. The campaign often raises the journalist’s ties to foreign entities and challenges her “Malagasy blood”.<sup>214</sup> People supporting her are called traitors.<sup>215</sup>

### 3.2.2.2 Case 2: Civil Society

During the 2018 presidential elections, civil society member and human rights defender Ketakandriana Rafitoson commented on a TV5Monde interview on the disproportions of means deployed by then-presidential candidate Rajoelina’s electoral campaign and raised suspicions of corruption on the account of the presidential candidate being sponsored by businessman Mamy Ravatomanga, an essential figure in Madagascar for investors, at the time being investigated before the Parquet National Financier (PNF) in France for “rosewood trafficking, corruption and tax evasion”.<sup>216</sup> The interview garnered its share of attention and outrage. Offline, Mrs. Rafitoson was threatened with defamation suits and interferences in her professional position. Though these threats were not followed upon, the smear campaign that spread in parallel on Facebook “had their effect” according to an interviewee.<sup>217</sup>

The main point of language used during the campaign accused Mrs. Rafitoson of being affiliated with Rajoelina’s political opponents and spreading defamatory information to harm his

---

<sup>208</sup> “Joint declaration,” 28 June 2021.

<sup>209</sup> Data 4, 7, 8, 10. “Global Leaders.”

<sup>210</sup> Data 7.

<sup>211</sup> Data 4, 7, 8, 10.

<sup>212</sup> Data 3.

<sup>213</sup> Data 6, where it is implied RSF, who published a statement in support of Borgia, works itself for French interests, referencing the territory dispute between Madagascar and the French Republic over the Scattered Islands. For more, see: Rakotoarivelo, “The dispute between Madagascar and France.”

<sup>214</sup> Data 4, 8.

<sup>215</sup> Data 7.

<sup>216</sup> Jeune Afrique, “Levée des saisies.”

<sup>217</sup> Interview 2.

reputation. Along with this, violent misogynistic and sexist language was used. Trolls reposted photos of her and her family members on Facebook and turned them into virulent memes. A difficulty in determining state responsibility for this case arises from the fact that the campaign happened at a time where Rajoelina was not yet in office, and the teams presupposed coordinating the smear campaigns on Facebook were not attached nor affiliated to the state. However, online attacks against Mrs. Rafitoson have since sporadically re-appeared with similar language strategies. Additionally, in the wake of Mrs. Rafitoson's interview, the Malagasy newspaper Free News published an article revealing personal information about her husband (doxxing<sup>218</sup>, as noted by Nyst and Monaco, is often a method used during state-trolling campaigns). The then editor in chief of Free News has since become Minister of Communications under Rajoelina's presidency, and is at the heart of the accusations relayed in the 2021 France24 report on rolls farms in Madagascar.

### 3.2.2.3 Case 3: Journalist

The covid-19 crisis in Madagascar brought upon a tightening of the public debate, with a ban on live radio interventions and the obligation for all private audiovisual media to broadcast state-produced programs devoted to the pandemic. The Covid Organics (CVO), a traditional remedy based on artemisia, was at the heart of the governmental response. On 20 April 2020, president Rajoelina announced in a television broadcast that his country had found a "preventive and curative" cure for COVID-19.<sup>219</sup> He publicly consumed the remedy and ordered a nation-wide distribution to families.<sup>220</sup> The CVO was developed and produced in Madagascar by the Malagasy Institute of Applied Research. Its launch garnered a wide range of scientific skepticism from within and outside Africa<sup>221</sup> but also raised queries within Malagasy society.

In July 2020, journalists investigating on the composition of the CVO and its supply chain production faced online attacks after interviewing the director of the Malagasy Institute of

---

<sup>218</sup> Searching for and publish private or identifying information about (a particular individual) on the internet, typically with malicious intent.

<sup>219</sup> Ventures Africa, "Madagascar Launches"; Africanews, "COVID-organics"; Premium Times, "Madagascar launches."

<sup>220</sup> The Guardian, "Madagascar hands out 'miracle'."

<sup>221</sup> Before cooperating with Madagascar, WHO issued a warning against use of an untested COVID-19 remedy. Later, the Covid-Organics was registered for WHO's Solidarity Trials. BBC News, "Do not use untested"; The Irish Times, "WHO skeptical." Science | AAAS, "Unproven herbal remedy"; L'express de Madagascar (in French), "Injection contre le conronavirus."

Applied Research. In a news report on July 30, the latter accused the journalists of fraudulent practices (that they introduced themselves with false pretense and documents). Later, personal information about the journalists (names and media affiliations) was leaked in a Facebook group, spurring a wave of online harassment<sup>222</sup>. Civil society collaborators of the journalists, by way of association, also faced online attacks. As in the two first cases, attacks accuse them of being a “pawn”<sup>223</sup> at the hands of foreign powers. A post by an account called Vaovao Androany accused the civil society organization collaborating with the journalist of willfully creating controversy to destabilize the country, saying, “Their real mission is to overthrow the government.”<sup>224</sup>

#### 3.2.2.4 Case 4: Opposition

The “opposition” differs to other cases in this selection as it does not exactly qualify as an “individually targeted disinformation campaign”, but rather a cluster of individually targeted campaigns presented as one case. Through our mapping process, we have found that these campaigns are less trigger-oriented and rather follow a recurring, systematic basis: a member of the opposition party calls a rally or comments on political affairs, and soon enough vitriolic online commentaries follow. It appears the target here, rather than the individual, is the entity they represent, which the campaigns paint as a constant adversary to progress and development. The campaigns often use disparaging comments on the physical and mental health of their targets, alluding to depression, senility, and other afflictions. They accuse targets of collusion with foreign interests and sabotage to destabilize and overthrow the government such as manipulating student strikes. People perceived as supporting the opposition are also accused of sabotage, extremism and fanaticism.

As an illustration, Hanitra Razafimanantsoa is an elected Member of Parliament from the main opposition party Tiako I Madagasikara (TIM). She is known to be an outspoken critic of the current government and serves as one of the leaders of the “Miara-Manonja” movement, an opposition media initiative producing radio and video episodes. Mrs. Razafimanantsoa has been targeted by online smear campaigns, accused of being “the instrument of Marc Raval-

---

<sup>222</sup> Data 51.

<sup>223</sup> *Id.*

<sup>224</sup> Data 55.



manana<sup>225</sup> and certain foreign powers to try to destabilize the regime.”<sup>226</sup> The attacks appear as a systematic reaction to her every moves and declarations. The campaigns accuse her of being ”radicalized”, of ”extremism”, “hysteria“ and “fanaticissm“ and question her mental health by pointing to ”disjointed and delirious remarks”, suicidal tendencies and paranoia.<sup>227</sup> In February 2021, the attacks extended to a humiliation campaign against her daughter, after the latter was arrested for possession of cannabis. The Bar Association of Madagascar speaks of an unjust conviction and an illegal investigation procedure, while Mrs. Razafimanantsoa denounces a political machination.<sup>228</sup>

### 3.2.2.5 Case 5: Whistleblower

In November 2020, Ravo Ramasomanana resigned from Madagascar's public health ministry. Working in the procurement management unit, he had refused to sign documents which he said were tainted with irregularities and demonstrated acts of corruption. He denounced these acts in a video, with supporting evidence, posted on Facebook in April 2021. His revelations concern 130 million Ariary (€285,689) spent on white paint intended for the facades of public hospitals, while under covid-10 hospitals severely lacked medicines, oxygen, protective materials and personnel, and suspicious movements of the Ministry of Health's funds amounting to 44 billion Ariary. Mr. Ramasomanana was summoned by the cybercrime cell and deferred to the courts, facing up to 5 years of imprisonment for threatening public order and inciting political unrest.<sup>229</sup> He was later acquitted on the benefit of the doubt but sentenced to pay a fine for “insults and defamation”, as per the law on cybercrime.<sup>230</sup>

Ramasomanana was supported by human rights entities and anti-corruption NGOs in Madagascar, including Amnesty International in Madagascar and Transparency International. He was also submitted to an online smear campaign, as the same post accusing him of being a paid mercenary of the opposition to spread disinformation was shared multiple times in multiple Facebook mega groups (8 groups ranging from 5000 to 40 000 subscribers, including 3 groups named after popular television channels in Madagascar, with their logo as profile pic-

---

<sup>225</sup> Former president of Madagascar ousted after a coup lead by Rajoelina in 2009, and leader of the TIM party.

<sup>226</sup> Data 14.

<sup>227</sup> Data 16.

<sup>228</sup> L'Express de Madagascar, “Procès – Maître Francesca. “ (In French)

<sup>229</sup> Amnesty International, “Madagascar: les autorités.“ (In French)

<sup>230</sup> RFI, “Madagascar: verdict.” (In French)

ture) in May 2021.<sup>231</sup> Other posts accused him of being a drug addict, doxxed and weaponized information on his mental health, and accused him of falsifying facts about his previous employment at the Ministry (“impersonating a public servant”).<sup>232</sup>

### 3.2.2.6 Case 6: Africa-France Summit

In 2021, a delegation of Malagasy activists attended the New Africa-France Summit, in Montpellier, France. The Summit purported to re-define the fundamentals of the relationship between France and the African continent through a novel format, as it brought into dialogue French president Emmanuel Macron with members of African youth and civil society instead of a meeting between Heads of State and Governments. The Summit has been criticized as a communication campaign to reaffirm the African policy carried by President Macron.<sup>233</sup> The absence of political peers during this meeting caused astonishment and irritation among Macron’s African counterparts.<sup>234</sup>

Members of the Malagasy delegation have faced vitriolic online reactions during and after the summit, being called “traitors” and “sold”, and accused of colluding with the French government and French enterprises against Madagascar’s interests. The core content of these posts, according to one of our interviewees, could be summed by the phrase “France has organized a Summit for the opposition.”<sup>235</sup> This phrase was twitted verbatim by the then Minister of Foreign Affairs<sup>236</sup>, in what can be analyzed against our state-trolling model as a form of dog whistling to indicate governmental support of the campaigns and their content. One interviewee confessed the smear campaign eclipsed the messages the civil society wanted to convey through their participation to this Summit, flawed as it might be.<sup>237</sup> So much so that an alliance of civil society organizations later published a joint press release, with the following disclaimer: “The members of the delegation wish to recall that militants and/or activists are

---

<sup>231</sup> Data 52.

<sup>232</sup> Data 53.

<sup>233</sup> IFRI,” Nouveau Sommet Afrique-France.”

<sup>234</sup> *Id.*

<sup>235</sup> Interview 1.

<sup>236</sup> Patrick Rajoelina (Twitter), “Sommet”.

<sup>237</sup> Interview 1.

not systematically assimilated to political "opponents". Civil society is neutral and non-partisan."<sup>238</sup>

### 3.2.3 General Findings

As discussed above and in 1.5, we have collected singular posts from the Facebook platform and through a process of pattern recognition guided by the model anatomy a state-trolling campaign<sup>239</sup> we aggregated them into concerted and coherent disinformation campaigns to illustrate state-sponsored disinformation in Madagascar. In this subsection, we discuss findings which apply across our selection.

First, campaign targets are generally politically outspoken people of "higher profile".<sup>240</sup> Those are people that would be labeled as "personalities" with some modicum of platform and audience whereas regular people speaking up against or in ways that would tarnish the image of the government, or its proxies face more direct consequences.<sup>241</sup> Targets can be journalists, whistleblowers, political commentators, or a person expressing a potentially dissenting view on Facebook, but in most cases, they are active members of civil society (or activists) or active/perceived members of the political opposition.<sup>242</sup>

Second, triggering events vary in location (online or offline), but all concern instances where the target expressed views which can be perceived as "dissenting" or revealed information which can be analyzed as painting the government in a negative light. Emphasis is put on "perceived" and "analyzed as" here as the dissenting or critical intent of the targeted speech is not always obvious. In Case 1, for example, the journalist showcasing the South's precarious famine victims alludes to weak governance, whilst in Case 6, civil society participation in the Africa-France summit offers a less pronounced link to governmental image or interest.

Third, the language used in the posts we studied follows the language points raised by Nyst and Monaco. The most recurring theme across cases is the accusation of fake news fabrication

---

<sup>238</sup> TI-MG (Facebook.com), "La delegation malgache."

<sup>239</sup> Woolley, Monaco & Nyst, "State-sponsored trolling." See above, 3.1.1.

<sup>240</sup> Interview 2.

<sup>241</sup> *Id.* Such consequences involve legal sanctions, arrests or police convocations, often on the premises of fake news dissemination.

<sup>242</sup> Mostly, members of the RMDM movement or allies and followers of former president Ravalomanana, and in some cases allies of former president Rajaonarimampianina.

and dissemination. According to Nyst and Monaco, trolls often accuse targets of the very behaviors the state is engaging in.<sup>243</sup> An adjacent recurring theme is campaigns portraying targets as working to destabilize the government for the benefit of often unspecified foreign entities: foreign donors, partners<sup>244</sup>, states,<sup>245</sup> or foreign press. People expressing supports for targets are labeled “traitors”.<sup>246</sup> These reflect two language tactics noted by Nyst and Monaco: accusation of collusion with foreign intelligence (here accusation of collusion with foreign interests) and accusations of treason. We also note the creation and use of elaborate cartoons and memes (for example, the target in Case 1 has been “memefied” as a “pseudo journalist”, with all accomplishment susceptible to support her claim to professionalism or integrity ironized into self-aggrandization). The use of violent hate speech as a means of overwhelming and intimidating targets varies across cases, but the tone of language is generally bellicose and vitriolic, relying on insults and mockery. Misogynistic language (death and rape threats, disparaging comments)<sup>247</sup> is common toward female targets.<sup>248</sup>

Whereas these first three criteria can be observed in variations across our selection, the two remaining ones (use of bots and algorithms; election antecedents) manifest in the same way in all our cases. These are also the two components in the model which cannot be directly observed our dedicated empirical methods (discourse analysis, social media scraping and interviews), which require for us to use proxy techniques (literature review, analysis).

This research does not include large data analysis or social bot detection techniques, which are necessary for a full answer on the use of bots and algorithms in a case<sup>249</sup>. However, our interview process did provide data confirming the existence of teams using Facebook algorithms to track online mentions of the Malagasy president or his close advisors.<sup>250</sup> We acknowledge the limitations in terms of rigor and reliability given this is based on perception data. An alternative, or additional, method would be to confirm the Facebook accounts sharing the aforementioned language points as fake accounts.

---

<sup>243</sup> Woolley, Monaco & Nyst, “State-sponsored trolling,” 13.

<sup>244</sup> Data 3.

<sup>245</sup> Data 6.

<sup>246</sup> Data 7.

<sup>247</sup> *Id.*

<sup>248</sup> Interview 1.

<sup>249</sup> See Karataş & Şahin, “A Review on Social Bot.”

<sup>250</sup> Interview 1.

In ICT and Artificial Intelligence (AI) studies, fake accounts are called “Sybils”. Sybils represent fake identities controlled by a small number of real users, used in coordinated campaigns to spread spam and malware.<sup>251</sup> The France24 report refers to fake Facebook accounts in the sense that they are created and operated by human beings posing as other people, rather than in the sense that they are operated by bots acting as automated social actors. Malagasy people online often claim they can intuitively recognize a fake account, or “compte fako” (a pun substituting the English word “fake” with the Malagasy word “fako” meaning “trash”). Among fake account identifiers, one interviewee cites: the absence of a profile picture (or a stolen one), few friends, few interactions and repetitive content in comments and posts.<sup>252</sup>

Karatas and Sahin distinguish structure-based detection techniques of Sybil accounts.<sup>253</sup> The fundamental assumption underlying this technique is that social networks show a “homophily tendency”.<sup>254</sup> That is, two connected accounts on a social network tend to have similar attributes. This assumption therefore grounds the “intuition” people have when detecting a possible fake account. Through our own observations, we found several Facebook accounts fitting the above “intuitive” description a fake account (no photos, no friends, no content or interaction other than sharing governmental realizations and decrying the untrustworthiness or dishonesty of the opposition or a critical individual). Often, these accounts are connected in some way: they comment on the same posts, share the same links, follow the same groups, etc. So though the actual use of bots and algorithms may remain a black box to us, the data and analysis at our disposition do point to a plausible use of it.

Finally, on the use of structures and resources gathered during election campaigns, the use of online attacks by the current governing party has been reported since the 2018 elections.<sup>255</sup> Case 2, notably, took place during that electoral period. One interviewee notes that 2018 is the time where Facebook “trolls” have started polluting Facebook groups dedicated to national and local news.<sup>256</sup> If the governing party started to resort to online disinformation tactics during the election times, it is probable that they continued since being in office.<sup>257</sup> In the

---

<sup>251</sup> Wang and others, “Social Turing Tests,” 1.

<sup>252</sup> Interview 1.

<sup>253</sup> Karataş & Şahin, “A Review on Social Bot,” 3.

<sup>254</sup> Gong, Frank and Mittal, “Sybilbelief,” 976.

<sup>255</sup> RFI, “Présidentielle Madagascar.”

<sup>256</sup> Interview 2.

<sup>257</sup> Interview 1.

France24 report, when confronted with such accusations, the Ministry of Communication responded that like in other countries, this government has supporters freely expressing political support through legal means.<sup>258</sup> One interviewee noted that though the government and its proxies are not the only entities engaging in online disinformation for political gain, they are better-placed in terms of resources and opportunities to profit from such tactics.<sup>259</sup>

For reading purposes, the six cases constituting our empirical mapping process, which we have presented and discussed above, are synthesized into Table 3.1.

---

<sup>258</sup> France 24, “Facebook 'troll farms'.”

<sup>259</sup> Interview 3.

**Table 3.1: Individually Targeted Disinformation Cases in Madagascar**

CRITERION	CASE 1	CASE 2	CASE 3	CASE 4	CASE 5	CASE 6
<b>Target</b>	Journalist	Civil society member	Activist	Opposition	Whistleblower	Journalist
<b>Trigger</b>	Video post showing the precariousness of the victims of a famine in the south of Madagascar	News channel interview during which the target commented on then President-to-be's electoral campaign funds origine	Malagasy delegation of activists and civil society member participating to the 2021 Montpellier Africa-France Summit	Major figure of opposition; attacks more systematic and more target oriented than trigger oriented	Posted a video on Facebook to disclose cases of high corruption in the Ministry of Health	Conduct an investigative report on the supply chain surrounding the fabrication of the COVID organics (CVO) which has been produced and promoted by the government
<b>Language</b>	Accusation of collusion with foreign interests; accusation of treason; accusation of fake news dissemination; discrediting memes and images; use of insults	Accusation of collusion with the opposition, use of violent hate speech, accusation of fabricating and spreading false information, infringement of privacy	Accusation of collusion with France and French enterprises; accusation of treason; use of misogynistic speech; amalgamation with the opposition	Accusation of collusion with foreign powers and interests; accusation of treason; accusation of sabotage; accusation of staging public unrest; use of memes and images	Accusation of forging fake news; Accusation of being paid by the opposition to pose as a public officer and stimulate public unrest; Disparaging comments on mental health	Accusation of fake news dissemination; accusation of sabotage and collusion with the opposition; use of discrediting insults and images
<b>Bots and algorithms</b>	<p>Use of algorithms to track posts mentioning the President or a close advisor</p> <p>Use of Sybils (fake accounts) to spread partisan speech while discrediting dissenting voices</p>					
<b>Election antecedents</b>	<p>Observed cases of online campaigns using fake accounts and spreading disruptive speech for political gain observed during presidential electoral period in 2018</p> <p>Current government most well placed in terms of resources and opportunities to profit from online campaigning tactics</p>					
<b>State Responsibility Mechanism</b>	State incited or fueled Statement of the local governor and counter-report with forged testimonies on national TV channel, accusing the journalist of spreading fake news (use of state proxy)	State coordinated - state incited Not formally the State at the time, but the revelations directly concern current president and his advisor	State incited or fueled Ministry of Foreign Affairs posted a tweet implying collusion with France and the opposition (dog whistling)	State incited Less conducive data	State incited	State incited or fueled Doxxing of information about the journalist after document shared on TV by a public medical institution director (state proxy)

### 3.3 Impacts analysis and Facebook's involvement

As we have outlined the “anatomy” of state-involving online disinformation campaigns in Madagascar, we now need to determine Facebook's level of involvement in the case. As we have seen, Facebook's content-grounded responsibility (see 2.1.2) and its para-business relationship to its users (see 2.2.2) make determining Facebook's involvement more challenging. For this reason, we have developed the Responsibility Scale in 2.3 to formalize scenarios which may engage the company's responsibility, which we will apply to our findings.

To determine Facebook's level of involvement, we need to determine whether the facts discussed in this chapter constitute a “severe negative impact” of the company's activities (as it has been established in 2.1.1 that freedom of expression constitutes a salient human right for Facebook, given the company's particular business footprint) that should be addressed and mitigated.

The GP's framework aims to identify where a company should focus its attention based on salience. Salience is determined first based on the severity of harm the risk poses to people.<sup>260</sup> Severity is determined by the harm's scale, scope and remediability.<sup>261</sup> For Madagascar, the adverse effects to right to freedom of expression appear to take place on a two-tier level of individual and societal effects.

On an individual level, these impacts are:

- an emotional toll of being subject to what has been described as nearing online harassment;
- collateral toll suffered by affiliates or close relationships (for example, one of the interviewee recounted how their niece was refused an administrative post because of the interviewee's public stands on political affairs);<sup>262</sup>
- privacy infringements (photo reposting, private information shared online);<sup>263</sup>

---

<sup>260</sup> UNHRC HR/PUB/12/02, 37.

<sup>261</sup> “Higg Brand & Retail Module Guidance.”

<sup>262</sup> Interview 1.



- self-censorship of users who would otherwise take part in the public debate, resulting from fear or apprehension of knowing that “thousands of bots and eyes may be spying on you across the digital landscape”,<sup>264</sup>
- self-censure of journalists;<sup>265</sup>
- loss of credibility: targets become isolated in the public debate;<sup>266</sup>

On a societal level, the impacts are:

- erosion of public trust in otherwise trustworthy sources;
- civil society actors are drowned out by waves of virality and performative outrage, and isolated due to the erosion of public trust, resulting in a weakened civil society;
- impacts on right to information, as journalists and contradictory voices resolve to self-censorship and as public debate and information diversity deteriorate;<sup>267</sup>
- such campaigns also displace the public debate on other subjects which do not criticize state actions, as a sort of diversion tactic;<sup>268</sup>

The second component to salience is likelihood. Likelihood is predicting how likely the impact may exist or occur in the company’s operations and supply chain. Although precedent Facebook Human Rights Impact Assessments (HRIA) will be discussed more in details later in Chapter 4, it is to be noted that Facebook has already had to contend with its platform being misused to spread hate speech and disinformation in countries such as Myanmar<sup>269</sup> and Sri Lanka.<sup>270</sup> Although no HRIsAs or other formal investigation have been conducted by the company regarding state-sponsored smear campaigns in particular, these cases have been extensively reported by stakeholders such as NGOs, activists, civil society and journalists.<sup>271</sup>

---

<sup>263</sup> Interview 1, 2.

<sup>264</sup> Interview 1.

<sup>265</sup> Interview 3.

<sup>266</sup> Interview 3.

<sup>267</sup> Interview 1.

<sup>268</sup> Interview 1.

<sup>269</sup> Quartz, “Facebook enabled.”

<sup>270</sup> The New York Times, “Where Countries.”

<sup>271</sup> See Introduction. For example, Maria Ressa accusing Facebook of either “negligence” or “complicity”.

Measuring likelihood also involves considering the countries the company is operating in, and the capacity of the company's business partners to manage human rights risks.<sup>272</sup> In 2.2, we argued that social media users resemble business relationships in the sense of the UNGP. Given the low level of instruction in Madagascar and the freedom of tones and lack of deontological imperatives on Facebook, Malagasy users are at risk of being more vulnerable and permeable to disinformation.<sup>273</sup> Moreover, the media landscape in Madagascar and the restrictions on expression result in a weakened information quality, to the benefit of bellicose partisan speech, "leaving less space for the dispassionate discourse needed for democratic debate."<sup>274</sup>

Thus, the adverse individual impacts observed in our case are compounded by severe societal impacts, which accentuate the already challenged free speech environment in Madagascar. Moreover, Facebook's previous experience with similar misuse of its platform should have enlightened the company on the risk of its platforms being weaponized. From this, we gather that Facebook's platform, as a product, and Facebook users, as business relationships of the company, have been instrumental in spreading disinformation content which reduced the ability of individual people in Madagascar to fully enjoy their right to freedom of expression. Referring to our Responsibility Scale, this situation corresponds to the "Linkage" level of involvement.

This provides us with the first part of the answer to our main research question, the second part of which will be addressed in the next chapter as we assess whether Facebook has provided the requisite effort to address its involvement in this case.

---

<sup>272</sup> "Higg Brand & Retail Module Guidance."

<sup>273</sup> Interview 1.

<sup>274</sup> Interview 3.

**Table 3.2: Facebook Responsibility Scale - Degree of Involvement**

Adapted from the UNGP's three ways companies can be involved with human rights impacts—cause, contribution and linkage--which define how companies are expected to respond. Designed for cases of free speech infringements involving the Facebook platform's user content.

Level of involvement	Non-impacting	Non-involved	Caused	Contributed	Linked	Complicit - non-legal	Complicit - legal
<b>General definition</b>	No human rights impact	Human rights impacts registered in the periphery of the company's activities but not engaging the company's responsibility under Pillar 2	Causes an impact through its own activities	Contributes to an impact either directly or through some outside entity (government, business or other)	A company's operations, products or services are linked to a negative human rights impact through a business relationship (or series of relationships)	Indirect involvement in human rights abuses	Aiding and abetting to human rights impacts National criminal liability or International criminal law jurisprudence: "knowingly providing practical assistance or encouragement that has a substantial effect on the commission of a crime".
<b>Facebook-specific definition</b>	Facebook's activities involving its user content do not infringe on the right to freedom of expression of others	Facebook's activities involving its user content restricts freedom of expression in conformity with international human rights law standards (Article 19(3) and 20 ICCPR)	Facebook directly infringes on freedom of expression with regard to user content	Facebook contributes to free speech impacts directly or through an other entity	Facebook is directly linked to a free speech infringement through the service its provide and its users' action	Facebook indirectly and knowingly causes/enables abuses to free speech by a third party	Facebook is aiding and abetting to unlawful restrictions on freedom of expression
<b>Scenario</b>	Facebook platforms posts in a way and content that does not infringe on freedom of expression	Facebook applies content moderation policies with respect to the demands of freedom of expression, and with satisfactory oversight and transparency	Facebook applies content moderation policy which generally targets and suppresses posts on certain subjects or by certain groups, in opposition with international standards for freedom of expression	Facebook contributes to harm to free speech by enabling harmful posts (providing the technical solution or failing to acknowledge and prevent harmful use of its platform)	Facebook users coordinate harmful posts on the company's platform without Facebook leveraging against it	Facebook allows/facilitates its platform to be used in harmful ways by a third party	Facebook allows/facilitates its platform to be used in unlawful and harmful ways by a third party
<b>Appropriate actions</b>	None	None	<ul style="list-style-type: none"> <li>• Cease the activity that caused the impact</li> <li>• Provide remedy</li> <li>• Take steps to prevent impact from recurring</li> </ul>	<ul style="list-style-type: none"> <li>• Cease activity and avoid contribution</li> <li>• Provide remedy</li> <li>• Use leverage to mitigate any remaining impact to the greatest extent possible</li> </ul>	<ul style="list-style-type: none"> <li>• Has forward-looking responsibility to prevent the impact from recurring</li> <li>• No explicit responsibility to provide remedy but can choose to do so</li> </ul>	-	-

## 4 Facebook's responses: in Madagascar, and beyond

This chapter reviews what Facebook does or has done to address the issue at hand (Sub-question 2), to see if the company provides the requisite effort and reasonable care to avoid harm to other persons to uphold their corporate responsibility in this case and where improvement remains necessary (Sub-question 3).

### 4.1 An array of untapped measures

#### 4.1.1 Case-specific measures, or lack thereof

As we have seen in 3.1.2, the Facebook platform has been instrumental to spreading state-coordinated and state-fueled disinformation campaigns in Madagascar which resulted in infringements to the right to freedom of expression of the people targeted by such campaigns. Pillar 2 of the UNGPs provides businesses should seek to prevent or mitigate adverse impacts they are linked to, and in the case of Facebook such impacts extend to harm involving user content (see 2.1.2). To uphold its corporate responsibility Facebook should address the cases happening in Madagascar to mitigate their impact (by exercising leverage) and to prevent it from recurring (by undertaking HRDD).

Referring to the cases we studied, Facebook should for example take measures to identify and take down harmful and defamatory campaigns such as the ones accusing a critical journalist of being a “traitor” (Case 1, Case 3) or the ones resorting to violent misogynistic speech to attack an outspoken civil society member (Case 2, Case 3, Case 6). Regarding the societal impact of these campaigns, Facebook should take measures to contain the spread of these campaigns and engage in building Malagasy users' capacity to responsibly engage with online content and potential disinformation. Our interviewees underlined the provision of a modicum of “media and information literacy” as part of the company's responsibilities.<sup>275</sup>

There is a broad range of practical measures that Facebook could adopt to address its responsibility in the Madagascar case, as the GPs are designed as evolutionary and adaptable, and offer guidance rather than injunction. However, there is little evidence of measures the com-

---

<sup>275</sup> Interview 1, 2 and 3.

pany has taken specifically in relation to Madagascar. One of our interviewees, for example, recounted instances in which campaign targets or their close ones have reported accounts and posts they deemed harmful and did not receive any follow up.<sup>276</sup> However, Facebook’s operations are global and cases of state-sponsored disinformation campaigns to silence dissenting voices are emerging worldwide as a key free speech issue (see 1.1). For these reasons, based on Facebook’s failure to respond to this phenomenon in Madagascar, we will look at the company’s general measures regarding harmful disinformation, as well as specific measures taken in cases similar to that of Madagascar, to guide our assessment. As there is little to assess regarding the question “what has been done specifically,” the next step is to consider “what has been done generally” along with “what could have been done specifically.”

Facebook has a wide array of company policies and measures to address harmful misuse of their online platforms, which we review below.

#### 4.1.2 Corporate Human Rights Policy

The UNGPs afford a central role to HRDD in the corporate responsibility to respect. The process should enable any corporation to achieve full respect for all human rights. A pre-requisite in undertaking HRDD is for the company to adopt and publish a high-level policy commitment to respect human rights.<sup>277</sup> This statement should reflect board-level acceptance to secure its proper implementation and should be public, so that external stakeholders have a clear platform to scrutinize companies affecting them.<sup>278</sup>

Facebook published its first corporate human rights policy in March 2021, affirming its commitment to the UNGPs in one central repository that applies to all its products, in all the geographies of its operation<sup>279</sup> (therefore, including both axes of our case: the Facebook platform as a product and Madagascar as an area of operation). This is done by cataloguing various policies, guidelines, principles and practices in one place, stating that Facebook’s respect for human rights is applied through their Community Standards<sup>280</sup>, their membership in the Glob-

---

<sup>276</sup> Interview 1.

<sup>277</sup> *UNHRC A/HRC/17/31*, Principle 15.

<sup>278</sup> O’Brien and Dhanarajan, “The Corporate Responsibility,” 8. *UNHRC A/HRC/17/31*, Principle 16.

<sup>279</sup> Meta, “Corporate Human Rights Policy.”

<sup>280</sup> Meta, “Community Standards.”

al Network Initiative<sup>281</sup>, the Responsible Supply Chain Program<sup>282</sup>, Privacy Principles<sup>283</sup>, Data Policy, Law Enforcement Guidelines<sup>284</sup>, Transparency Reporting<sup>285</sup>, Responsible AI efforts<sup>286</sup> and Diversity and Inclusion Practices<sup>287</sup>. The policy commits to measures such as upholding end-to-end encryption, protecting human rights defenders, providing remedy for human rights harms, and reporting on the policy’s implementation.<sup>288</sup> The policy answers to the UNGPs’ demand of “know and show”, as put by Miranda Sissons, Human Rights Director at Facebook: “Know the standards. Use the UNGPs. Be transparent. Show don’t tell.”<sup>289</sup>

Formalizing Facebook’s human rights commitment is a welcome step,<sup>290</sup> but the policy arrives after that of other ‘Big Tech’ companies such as Apple<sup>291</sup>, Google<sup>292</sup>, Microsoft<sup>293</sup> and Amazon<sup>294</sup>. Given Facebook’s market leadership since the 2000s and its past harms record, the policy might be “too little, too late”. The policy also suffers from its vagueness, and core issues remain unaddressed. For example, Facebook reiterates its commitment to its Community Standards and its recognition of the Oversight Board<sup>295</sup> but remains opaque on how it moderates content at the first instance. Similarly, Facebook commits to “report annually on how we’re addressing human rights impacts, including relevant insights arising from human rights due diligence, and the actions we are taking in response”<sup>296</sup> without specifying the particulars of these reports. As pointed by Ranking Digital Rights, “Facebook should significantly improve its transparency on and accountability for its content moderation by publishing consistent data on actions it takes to enforce platform rules.”<sup>297</sup>

---

<sup>281</sup> GNI, “Protecting and Advancing.”

<sup>282</sup> Meta, “Responsible Supply Chain.”

<sup>283</sup> Meta, “Privacy Principles.”

<sup>284</sup> Meta, “Information for Law.”

<sup>285</sup> Meta, “Transparency Reports.”

<sup>286</sup> Meta AI, “Collaboration on the future.”

<sup>287</sup> Facebook, “Diversity.”

<sup>288</sup> Meta, “Corporate Human Rights Policy.”

<sup>289</sup> OpinioJuris, “Facebook’s Commitment.”

<sup>290</sup> Accessnow, “After nearly two decades.”

<sup>291</sup> Apple, “Our Commitments.”

<sup>292</sup> Google, “Human Rights.”

<sup>293</sup> Microsoft, “Human Rights Statement.”

<sup>294</sup> Amazon, “Human Rights Principles.”

<sup>295</sup> The Verge, “Facebook Oversight Board.”

<sup>296</sup> Meta, “Corporate Human Rights Policy”; ADL, “Facebook’s transparency reporting.”

<sup>297</sup> Ranking Digital Rights, “Facebook, Inc.”

### 4.1.3 Due Diligence on Disinformation

Beyond the “commitment” prerequisite, due diligence as envisaged by the GPs<sup>298</sup> comprises four subsequent steps resembling a typical continuous improvement cycle.<sup>299</sup> These steps are:

- Assessing actual and potential impacts of business activities on human rights;
- Acting on the findings of this assessment, including by integrating appropriate measures to address impacts into company policies and practices;
- Tracking how effective the measures the company has taken are in preventing or mitigating adverse human rights impacts;
- Communicating about the due diligence process and its results.

#### 4.1.3.1 Assess

The administration of a HRIA is the first step in the GPs’ due diligence process. An adverse human rights impact may be said to occur when an action removes or reduces the ability of an individual to enjoy his or her human rights,<sup>300</sup> and thus the parameters and process of HRIA under the GPs remain emergent and contested.<sup>301</sup> In recognition of the procedural difficulties inherent with large corporations, Principle 17 of the UNGPs allows for corporations to “identify general areas where the risk of adverse human rights impacts is most significant [...] and prioritize these for human rights due diligence.”

Facebook has been conducting case-specific HRIAs before publishing its human rights statement. There has been to date no HRIA conducted in Madagascar, as discussed in 3.3 and 4.1.1, but previous HRIAs hold a modicum of relevance. As per the UNGP, their findings should have informed Facebook to take measures and integrate them within its management system and operations to alleviate future impacts. The previous HRIAs in question evaluate the role of Facebook’s services in Sri Lanka, Indonesia, Myanmar, and Cambodia (see Table 4.1).

---

<sup>298</sup> See 2.1.2.

<sup>299</sup> *UNHRC A/HRC/17/31*, Principles 17-20.

<sup>300</sup> IBLF, “Guide to Human Rights,” 24.

<sup>301</sup> O’Brien and Dhanarajan, “The Corporate Responsibility,” 11.

Points raised in these HRIAs echo aspects of the Madagascar case. The duality of Facebook use in Cambodia reminds the Malagasy context (see 3.1). As put by the HRIA: “while Facebook is widely used as a platform for free speech, many users also exercise a significant degree of self-censorship; while Facebook is an essential outlet for media organizations, it is also home to increasingly sophisticated disinformation efforts; while Facebook provides a place for civil society to organize, rights holders have been arrested for content they have posted, liked, or shared on Facebook.”<sup>302</sup> However, the facts analyzed in Madagascar are more akin to those observed in Indonesia, notably regarding “self-censorship of users” as a free speech impact, and the risks identified both on an individual and a societal level. Interestingly, for Myanmar, similar impacts to our discussions in 3.3 appear under risks related to security as and ICCPR 20, possibly as Myanmar involved hate speech intended to incite offline violence rather than partisan disinformation intended to control public debate.

Recommendations in all four HRIAs underline the need to invest in functional teams mastering the local context and language and more engagement of local stakeholders such as civil society (by consultation for transparency reporting, for example) and users (through digital literacy initiatives, for example). The recommendation to perform HRIAs in other at-risk countries and more systematically is also a recurring point. Indeed, beyond those HRIAs, there are documented cases of harmful social impacts due to Facebook’s conduct in different jurisdictions throughout the world. In 2020, Ranking Digital Rights has reprimanded Facebook on how it conducts HRDD and questioned whether Facebook “conducts systematic impact assessments of its terms of service enforcement, on its targeted advertising policies and practices, on its development and deployment of algorithmic systems, or on its zero rating programs.”<sup>303</sup> These recommended developments remain unimplemented in Madagascar, which shows Facebook’s failure to integrate its HRDD findings in its overall operations.

---

<sup>302</sup> BSR, “Cambodia,” 3.

<sup>303</sup> Ranking Digital Rights, “Facebook, Inc.”



**Table 4.1: Overview of independent HRIAs commissioned by Facebook**

Source: Article One, "Sri Lanka" & "Indonesia; BSR, "Myanmar" & "Cambodia."

Country	Context	Risks identified	Recommendations
Sri Lanka	Ethnic and religious tensions playing out both online and offline One of the most critical countries for potential human rights infringements on the Facebook platform	Risks related to vulnerable groups, including online harassment of HRDs by other users  Government overreach to monitor defender activity online, which infringes on free expression (UDHR 19) if resulting in self-censorship  Facebook’s response slow and insufficient to meet UNGP standard: Lack of formal human rights due diligence Limited cultural and language expertise among Facebook staff	Expand country-level HRIAs to other critical countries  Publish data on content moderators, including the number of content moderators disaggregated by language expertise, country of origin, age, gender, and location  Make Community Standards accessible and understandable, including by ensuring they are translated into all languages spoken by Facebook users in Sri Lanka
Indonesia	Ongoing domestic limitations on the right to free expression, especially targeted against non-Muslims  Disinformation and misinformation issues (“hoaxes”).	Societal risks resulting from disinformation and misinformation  “Hoax factories” used to target political figures and promote politically motivated disinformation in the hopes of impacting national and local elections	Formalization of an “At-Risk Country” list  Making Community Standards sufficiently accessible and understandable  Developing processes to apply cultural context to content moderation
Myanmar	Evaluated the use of Facebook to spread anti-Rohingya sentiment in Myanmar in 2018  UN Fact-Finding Mission on Myanmar concluded that serious crimes under international law have been committed by military and security forces that warrant criminal investigation and prosecution	Hate speech, incitement to violence, and content intended to coordinate harm may lead users to self-censorship as a potential free speech  Content intended to harass users and silence them in the public debate are found under risks related to security as per UDHR 3 and ICCPR 20	Forward-looking recommendations shaped by the observation that Facebook’s human rights impacts in Myanmar cannot be addressed by Facebook alone  Enforce community standards through a cross-functional team that understands the local context  Publish a local, Myanmar-specific version of Facebook’s Community Standards Enforcement Report  Develop a more robust and systematic approach to human rights by undertaking HRIAs in other high-risk markets
Cambodia	Facebook recognized by local stakeholders as an important space for freedom of information and expression, political participation, and government accountability  Challenging legal, social, and political context, with civic space being actively constrained and low digital literacy levels	Assessment prioritizes freedom of expression as a salient human right to be taken into consideration, although for different risks than the ones in our case	Apply leverage on the legal and regulatory framework in Cambodia as it relates to social media platforms  Publish a local Cambodia-specific version of the Community Standards Enforcement Report

#### 4.1.3.2 Act

Once an assessment is completed, the GPs call for businesses to respond to its findings, to prevent future abuses and address any existing ones that may have been uncovered.<sup>304</sup>

According to Article One, which was commissioned for the HRIAs in Sri Lanka and Indonesia<sup>305</sup>, Facebook implemented numerous recommendations they proposed, making significant strides in addressing the impacts of its products and respecting the rights of users in each country. Specific changes Facebook made in response to these recommendations included a new policy to remove verified harmful misinformation, the expansion of a bullying policy to increase protections to individuals beyond children, and operational management changes through launching the Third Party Fact Checkers and hiring a new Human Rights Director.<sup>306</sup>

As with the Corporate Human Rights Policy, these responses have been criticized for appearing appeasing and reactive rather than proactive.

According to Business for Social Responsibility (BSR), Facebook's response to the HRIA in Cambodia sets out the meaningful steps to implement the recommendations.<sup>307</sup> However, said responses overlook recommendations concerning active advocacy for regulatory change in Cambodia and other forms of support for human rights defenders and vulnerable groups, citing user security as a reason to "not discuss detailed action steps in public."<sup>308</sup> As for the HRIA in Myanmar, Facebook has dedicated teams to work on issues specific to Myanmar, and planned to grow the team of native Myanmar language speakers reviewing content to at least 100 by the end of 2018.<sup>309</sup> Nonetheless, Facebook sustains criticisms for not having done enough, given their societal reach, to prevent the weaponization of their platform in Myanmar.

These positive measures, however criticized, show that there was precedent that the company could and should have built upon to address their adverse impacts in countries presenting similar risks and challenges as those identified in Myanmar, Cambodia, Sri Lanka and Indonesia. As shown earlier, the cases studied in Madagascar present non-negligible similitudes with the

---

<sup>304</sup> *UNHRC A/HRC/17/31*, Principle 24.

<sup>305</sup> Article One, "Facebook."

<sup>306</sup> For a full list of actions taken, see Meta, "Facebook Response: Sri Lanka."

<sup>307</sup> For more details, see Meta, "Facebook Response: Cambodia."

<sup>308</sup> *Id.*, 9.

<sup>309</sup> Meta, "An independent Assessment."

cases in Cambodia and Indonesia in particular. The lack of responses in Madagascar thus appears all the less justified.

#### 4.1.3.3 *Track and Communicate*

The UNGPs enjoin companies to track the effectiveness of their human rights approach by drawing upon feedback from external sources, including affected stakeholders, and “communicate” on how they are addressing their human rights impacts.<sup>310</sup> BSR commends Facebook’s disclosures of the findings of the HRIAs it has commissioned including the actual and potential adverse human rights impacts identified and actions taken, as it remains rare for companies to disclose such findings.<sup>311</sup>

On the other hand, the company faces criticism for the delayed release of an India HRIA and the lack of action to addressing its role in spreading hate speech and incitement to violence on its services in India.<sup>312</sup> According to a consortium of civil society organizations in India: “The current perception is that Facebook is not committed to respecting rights in this case...The UN Guiding Principles on Business and Human Rights are clear that transparency is a key aspect of human rights due diligence and that to account for how they address human rights impacts, companies should be prepared to communicate this externally, particularly when concerns are raised by or on behalf of the affected stakeholders.”<sup>313</sup>

#### 4.1.4 Voluntary Measures on Disinformation

Beside its HRDD process, Facebook has taken voluntary measures to tackle its adverse impacts on freedom of expression linked to disinformation and online harassment, which could have been relevant to address the cases in Madagascar.

First, the **Independent Oversight Board**,<sup>314</sup> created in 2018 as an independent operational grievance mechanism. The Board provides an additional remedy to users and helps Facebook answer the difficult questions around freedom of expression online: what to take down, what to leave up, and why. In its first decisions, the Board demonstrated its independence in over-

---

<sup>310</sup> UNHRC A/HRC/17/31, Principle 20-21.

<sup>311</sup> BSR, “A Human Rights Impact.”

<sup>312</sup> Time, “Facebook’s Ties” and Jurist, “Civil Society Questions.”

<sup>313</sup> Id.

<sup>314</sup> Douek, “Facebook’s ‘Oversight Board’.”

ruling four of Facebook’s decisions. The Board elicited a variety of responses, with Klonick describing its creation as an historic endeavor,<sup>315</sup> and The Verge deeming it "a wild new experiment in platform governance".<sup>316</sup>

Second, the **Third-Party Fact-Checking Program** relies on independent third-party fact-checkers certified through the International Fact-Checking Network to identify, review and rate potential misinformation across Facebook.<sup>317</sup> The Program differs from Facebook’s moderation policies as fact-checkers do not remove content, accounts or pages from the Facebook platform. Instead, each time a fact-checker rates content as false, Facebook communicates the findings to users and reduces the content’s distribution. The Program was created in 2016, after Facebook received criticism about online disinformation during the US presidential election<sup>318</sup> The company discloses countries in which they have third-party fact-checkers but gives no insight to the scale of the initiative for each country (number, budget, etc.)<sup>319</sup>. There are no third-party checkers for Madagascar.<sup>320</sup> The Program proposes a novel approach but has been criticized for a lack of transparency on its impacts<sup>321</sup> which has spurred Facebook’s fact-checking partners to question whether this is one more crucial project is given the minimum enthusiasm necessary for public relations.<sup>322</sup>

## 4.2 An insufficient implementation

Despite the array of policies that Facebook applies to address its adverse impacts involving its platform, online coordinated attacks in Madagascar remain unaddressed. Beyond the Madagascar case, Facebook’s measures have been criticized as “image repair initiatives”<sup>323</sup> and demonstrate the company’s failure to apply due diligence consistently and adequately.

---

<sup>315</sup> Klonick, "The Facebook Oversight Board", 2418.

<sup>316</sup> The Verge, "Facebook's new Oversight Board."

<sup>317</sup> Meta, "Meta’s Third Party."

<sup>318</sup> CJR, "Facebook’s Fact-checking."

<sup>319</sup> Nobel Peace Center, "Is Facebook Undermining."

<sup>320</sup> Meta, "Where we have fact-checking."

<sup>321</sup> Full Fact, "Report on the Facebook," 31; BBC, "Matter of fact-checkers"; The Guardian, "They don’t care."

<sup>322</sup> CJR, "Facebook’s Fact-checking."

<sup>323</sup> Journalism Professor for Oslo Metropolitan University, asking Frances Haugen how she evaluates Facebook’s initiatives to appear more socially responsible. Nobel Peace Center, "Is Facebook Undermining."

#### 4.2.1 Deflecting Responsibility

Although none of the HRIAs developed above concern Madagascar, their findings should have informed Facebook enough to identify and anticipate the misuse of their platform to harass or discredit dissenting individuals as a risk beyond the single country level. Indeed, the UNGP provide that: “In order to prevent and mitigate adverse human rights impacts, business enterprises should integrate the findings from their impact assessments across relevant internal functions and processes and take appropriate action.”<sup>324</sup> As mentioned before, Facebook’s human rights statement applies globally, while the HRDD process mimics an improvement cycle based on a “horizontal integration across the business enterprise of specific findings.”<sup>325</sup> Theoretically, if the assessment findings have been given due weight, and acted upon, it should have provided Facebook with enough foresight to identify other at-risk countries and extent prevention and mitigation to these countries. Given the free speech challenges in Madagascar we discussed in 3.1, which as we have mentioned shares similarities with the Cambodian context, Madagascar fits among such countries.<sup>326</sup>

However, the commentary to Principle 19 recalls: “The horizontal integration across the business enterprise of specific findings from assessing human rights impacts can only be effective if its human rights policy commitment has been embedded into all relevant business functions.” Facebook’s human rights policy, for one, dates from little more than a year and remains to be operationalized. Hattotuwa also argues that Facebook needs to own the harm it contributed to when it did not have this policy.<sup>327</sup> The Facebook platform has contributed to ethno-political unrest, religious extremism and violent nationalism as far back as 2013<sup>328</sup> and Hattotuwa notes “how the company’s lip-service to rights didn’t sync with corporate imperatives”<sup>329</sup>. She recalls Facebook only started addressing some of these issues when impacts entrenched in Sri Lanka and other Global South markets<sup>330</sup> echoed at scale in the West. While releasing its HRIA on Sri Lanka, Facebook deplored the “misuse of our platform”, and

---

<sup>324</sup> *UNHRC A/HRC/17/31*, Principle 19

<sup>325</sup> *Id.*, incl. commentary.

<sup>326</sup> Moreover, Facebook already had to intervene to suspend three networks of Russian accounts that attempted to interfere in the domestic politics of eight African countries, including Madagascar. See Reuters, “Facebook says.”

<sup>327</sup> IHRB, “Making Facebook.”

<sup>328</sup> CPA, “Liking violence.”

<sup>329</sup> OpinioJuris, “Facebook’s Commitment.”

<sup>330</sup> WSJ, “Facebook Prepares.”

apologized for “the very real human rights impacts that resulted”. It did not however recognize that it had done enough to stop the spread of content inciting hate. Instead, it placed the blame on those who misused the platform although at the time there were no well-established guardrails against platform harms.

A close reading of the new human rights policy shows the problem persists. In Section 4, Facebook recognizes that “human rights defenders are a high-risk user group” but the same paragraph deflects and downplays risks HRDs face on Facebook’s products and platforms. The human rights policy names the company when there is a statement of corporate due diligence. However, when specific harms are noted, the blame shifts to “social media” at large. Even in its most progressive document on human rights, Facebook shies away from corporate responsibility of significant contributions to offline and online violence.<sup>331</sup>

#### 4.2.2 Understating Leverage Capacity

The commentary to UNGP 19 states that: “If the business enterprise has leverage to prevent or mitigate the adverse impact, it should exercise it. And if it lacks leverage there may be ways for the enterprise to increase it. Leverage may be increased by, for example, offering capacity-building or other incentives to the related entity, or collaborating with other actors.” Understandings of what constitutes leverage have been discussed earlier in 2.1.2 and 2.2.2.

Besides the above-mentioned account cleansing in 2019, no measures in Madagascar have emanated or are supported in a substantial manner by the company to suggest an intention to leverage the adverse impacts discussed in this research.

The commentary on the UNGPs recognizes that there are situations in which the enterprise lacks the leverage to prevent or mitigate adverse impacts. In their submission to the Special Rapporteur on Freedom of Opinion and Expression for her Report on Disinformation, Facebook refers to challenges in addressing disinformation campaigns on their platforms, mainly:

- the definitional challenge inherent to “disinformation” and other integrity issues;<sup>332</sup>
- the dilemmas of content moderation, as explored in Chapter 2 (who decides what is untruthful and what the penalties for untruthful content should be, as measures at-

---

<sup>331</sup> IHRB, “Making Facebook.”

<sup>332</sup> Facebook, “Facebook Submission,” 1.

tempting to address these questions risks undue restrictions on freedom of expression and the right to information);

- the highly contextual character of issues involving disinformation campaigns, such as language barriers, cultural context, legal ambiguity around the definition of disinformation;
- the challenge of identifying, removing and restricting access to harmful disinformation on a platform with billions of users;
- the complexity borne from the enormous scale of internet use worldwide, the rapid development of new technologies and increased political polarization in many countries.

These are legitimate challenges that a social media provider like Facebook is bound to encounter. However, given the scope of Facebook’s operations and worth, these challenges can hardly be considered enough to make leverage or mitigation impossible, even in a culturally and economically remote context like Madagascar. Facebook classifies among the ‘Big Five’ of tech companies along Alphabet (Google), Amazon, Apple and Microsoft.<sup>333</sup> These companies dominate their respective areas of technology, namely social media for Facebook. They have been among the most valuable public companies globally,<sup>334</sup> each having had a maximum market capitalization ranging from around \$1 trillion to above \$3 trillion<sup>335</sup>.

In 2021, Facebook has an estimated net worth of \$124.879 billion and market capitalization of \$ 939 billion, with 1.93 billion monthly active users as of December 2021,<sup>336</sup> a true Leviathan superseding the GDP of many countries including Switzerland, Saudi Arabia, and Madagascar.<sup>337</sup> This economic weight is compounded by the societal role Facebook plays in many countries. With more users than India and China’s populations combined, Facebook is the internet in many regions outside the West.<sup>338</sup> In many countries in Africa, the largest news

---

<sup>333</sup> Financial Times, “The Economics of Big Tech.”

<sup>334</sup> Reuters, “Apple becomes first company.”

<sup>335</sup> *Id.*

<sup>336</sup> GoBankinRates, “How Much Is Facebook.”

<sup>337</sup> Visual Capitalist, “The World’s Tech Giants.”

<sup>338</sup> IHRB, “Making Facebook”; Nobel Peace Center, “Is Facebook Undermining”; see 3.1.

sources are Facebook mega groups.<sup>339</sup> In 3.2, we brought up the case of a whistleblower who has sustained defamatory accusations across eight such groups. We have also underlined the role of Facebook regarding freedom of expression and access to information in the Malagasy context (see 3.1). Thus, along the denial of its full responsibility, Facebook downplays the large societal role it plays.

#### 4.2.3 North-South Disparities

Facebook's influence and the scope to which it facilitates harmful disinformation<sup>340</sup> is unprecedented, yet the company remains "chronically underfunded and understaffed".<sup>341</sup> Facebook has come to acknowledge to a degree the level of harm it is at risk of creating yet doesn't allocate resources accordingly. The Facebook Papers revealed the discrepancy between what Facebook monetizes in the United States (55 USD per person), in Europe (14.5 USD per person), in Asia (4.5 USD per person) and in the rest of the world combined (2.5 USD per person).<sup>342</sup> Haugen notes an unequal distribution of third-party fact checking program across countries of operations.<sup>343</sup> In 2020, 87% of the operational budget for misinformation went to English, even though only 9% of users spoke English.<sup>344</sup> This shows that Facebook consistently allocates attention for safety, not based on actual severity.

Irene Khan has remarked that "the largest [digital] companies, based in the United States and influenced by its politics and public opinion there, appear to be driven by United States and European priorities. They do not invest sufficient resources in understanding the local factors that feed disinformation online in other parts of the world, especially developing countries."<sup>345</sup> Facebook does not contradict this assessment. Facebook users outside of North America contract with the company's Irish subsidiary. This situation permits a divide in standards between operations in North America, where stronger regulation is compounded with more resources and attention, and the rest of the world, especially the Global South.<sup>346</sup> In 2021, Frances Haugen disclosed documents showing the deliberate choices Facebook makes

---

<sup>339</sup> Nobel Peace Center, "Is Facebook Undermining."

<sup>340</sup> For more on Facebook's enduring system of harm, see: WSJ, "Facebook Papers."

<sup>341</sup> Nobel Peace Center, "Is Facebook Undermining."

<sup>342</sup> *Id.*

<sup>343</sup> *Id.*

<sup>344</sup> *Id.*

<sup>345</sup> UNHRC A/HRC/47/25, para. 76.

<sup>346</sup> WSJ, "Facebook Papers."



to maximize profit at the expense of public safety, in what has come to known as “The Facebook Files.” The Files confirmed Facebook chronically underinvests in non-Western countries, leaving users exposed to disinformation, hate speech, and violent content.<sup>347</sup>

As seen in our Responsibility Scale, internet companies often risk linkage to human rights impacts that they do not cause or contribute. It may be argued that in this case, there is no direct or meaningful involvement of Facebook in the adverse impacts observed in Madagascar. Indeed, entities affiliated with the State are using the Facebook platform to propagate defamatory content to overwhelm, discredit and silence critical individuals. However, it could be said that such intents fall under fair use of the platform and should not engage the company’s responsibility, the same way a person shooting another person does not engage the responsibility of the gun manufacturer, though it may engage the responsibility of the dealer if the gun was knowingly sold to an intoxicated person<sup>348</sup> or a minor<sup>349</sup>, in a word if the dealer showed negligence, either by not making its due diligence or by ignoring its findings.<sup>350</sup> Interestingly, Facebook as company acts as both the manufacturer (Facebook designs the platform independently<sup>351</sup>) and the dealer (Facebook provides the platform to users, and lobbies for cheaper access prices in Global South countries<sup>352</sup>).

Following this logic, one could maintain that Facebook did not do anything in the case we are studying besides providing the service and product it is its business to provide, in respect with all relevant regulations. Facebook itself has argued this way when confronted with its involvement in cases such as Myanmar, deflecting responsibility to its users or to the State. However, we argue that the claim that “Facebook did nothing” is the very core basis of Facebook’s involvement and responsibility in Madagascar. Facebook published and designed an array of measures and initiatives which we have analyzed in 4.1. However, these measures did not translate enough into the company's operations to absolve their linkage to the free speech infringements evoked in this paper.

---

<sup>347</sup> See CNN Business, “Facebook knew” ; Rest of World, “The Facebook papers.”

<sup>348</sup> *Kitchen v. K-Mart Corp.*

<sup>349</sup> *Farley v. Snug Enterprises, Inc., et al.*; *Hoosier v. Randa.*

<sup>350</sup> To engage firearm manufacturers’ responsibility in cases of gun violence, the main criteria trend is foreseeability. The more foreseeable a shooting is, the more responsible the party who should foresee it.

<sup>351</sup> Benesch, “But Facebook’s”; Haugen, in Nobel Peace Center, “Is Facebook Undermining.”

<sup>352</sup> To attract new users, Facebook made deals with cellular carriers in countries including Pakistan, Indonesia and the Philippines to let low-income people use a limited version of Facebook without data charges. See WSJ, “Facebook Promised.”

## 4.3 Recommended measures

### 4.3.1 UNGP measures

To answer our last research question, many recommendations can be made to improve Facebook's responses in Madagascar and in general.<sup>353</sup> Based on our earlier analysis, these can be synthesized into two main themes: transparency and scale.

On scale, Facebook should extend its due diligence process to counter the geographical disparities in its policies' implementations by conducting more HRIAS in at-risk countries and less affluent markets. For our case, an HRIA in Madagascar is necessary, especially with the next presidential election scheduled for 2023. Facebook needs to adopt linguistic acuity and equity policies, for example ensuring their community standards are available in relevant languages in the countries where they operate and allocating adequate resources in at-risk countries such as third party checkers mastering the local language and context. According to Khan, "A thorough understanding of the local political, social and economic context, language proficiency and close cooperation with civil society in countries where disinformation is more prevalent are necessary."<sup>354</sup>

On transparency, beyond publishing regular updates on the implementation of its human right related policies, subject to stakeholder review,<sup>355</sup> Facebook should open its research to people outside the company. A lack of transparency regarding content moderation systems prevents adequate assessments of the effectiveness of the measures adopted by the company and its human rights impacts. Facebook data remains a "black box" for researchers and users.<sup>356</sup> Its transparency report, for example, provides information on the removal of fake accounts but not about content. Similarly, no information is provided about the ratio of disinformation content or the level of user engagement with disinformation, including numbers of shares, views, reach and number of complaints or requests for removal.<sup>357</sup> This confirms Benesch's observation that "internet companies govern, without meaningful accountability to the people under

---

<sup>353</sup> See for example *UNHRC A/HRC/47/25*, paras. 95-100.

<sup>354</sup> *UNHRC A/HRC/47/25*, para. 76

<sup>355</sup> Accessnow, "After nearly two decades."

<sup>356</sup> Nobel Peace Center, "Is Facebook Undermining?"

<sup>357</sup> *UNHRC A/HRC/47/25*, para 81.

their control.”<sup>358</sup> According to Haugen, “we can develop privacy-conscious ways of exposing radically more data than is available today, [...] It is important for our ability to understand how algorithms work, how Facebook shapes the information, to have these data sets publicly available for scrutiny.”<sup>359</sup>

#### 4.3.2 Beyond UNGP?

In her paper “But Facebook’s not a country”, Benesch argues that Facebook should be held to a higher standard of responsibility given the magnitude of its impacts: as a social media company, Facebook can limit “not only what people say (or write) but what they see and read”. When exercised in public spaces, such powers are reserved for governments. This extraordinary, transnational power sets companies like Facebook apart from other corporations, as it can influence the political, cultural, and economic development of entire societies.<sup>360</sup>

Moreover, the UNGPs’ limitations have often been criticized. Biltchiz deplores their inadequacy to address the exact nature and extent of the obligations and expectations they assign on States and corporations, while Deva deems the lack of positive human rights obligations binding corporations as a regressive approach.<sup>361</sup> It is thus important to consider an extended or more intensive regime of corporate responsibility for social media companies. Auto-regulation so far has borne lackluster results, at least in the case of Facebook, as Haugen hammers: “platforms cannot be trusted to act in the public interest.”<sup>362</sup>

One alternative lies within domestic and regional regulation. The incoming the EU Digital Services Act (DSA) and Digital Market Act (DMA), which set to strengthen rules around tech companies to create a safer digital space, promise an important advancement. However, there is no silver bullet in ensuring accountability for social media disinformation issues given the complex tangle of state and corporate responsibility, and countervailing rights, those entail. Alone, regulatory frameworks risk entertaining the rift in standards between countries having these regulations and countries not having them. When the EU applies the DSA, it will hopefully engineer an environment better suited to protect such rights as freedom of expression

---

<sup>358</sup> Benesch, “But Facebook’s”, 94.

<sup>359</sup> Engadget, “What Facebook.”

<sup>360</sup> BuzzFeed News, “I have blood on my hands.”

<sup>361</sup> Deva, *Regulating Corporate*, 109.

<sup>362</sup> Nobel Peace Center, “Is Facebook Undermining.”

online – in the EU. When the US Congress raises the issues of disinformation on 2016 Facebook, Facebook complies with the Oversight Board and a renewed commitment to truth-seeking – in the US. Whatever improvement the company adopts through these frameworks scarcely reaches more vulnerable regimes such as Madagascar, where local regulation remains challenging, as our interviewee expressed: “Today, I think that Facebook has become, somewhere, a ‘weapon’ in the hand of the state [...] and the state is not going to repress itself, it is not going to restrain itself.”<sup>363</sup>

De Schutter argues that only an international treaty could establish uniform corporate accountability irrespective of whether states comply with their duty to protect or not.<sup>364</sup> On social media businesses, Benesch argues that adhering to a single body of standards would give companies a source of “forceful normative responses against undue State restrictions” and would provide a basis for users to hold companies accountable.<sup>365</sup> In our introduction, we mentioned the ongoing drafting process of a “Legally Binding Instrument to Regulate, in International Human Rights Law, the Activities of Transnational Corporations and Other Business Enterprises.” A similar, sector specific initiative, might offer the single body of standards needed to counter social media businesses’ uneven enforcement of their own community standards. Similarly, Haugen advocates for a dedicated “oversight body” with the power to oversee social media platforms.<sup>366</sup>

Controversy remains regarding best path towards business respect for human rights: a binding instrument or a practical approach focused on national law and voluntary initiatives?<sup>367</sup> Ruggie maintains that a binding instrument offers more difficulties than advantages<sup>368</sup> while Deva considers “single instrument” or “single strategy” approaches misguided and emphasizes the need for complementary approaches coordinating persuasion with sanctions,<sup>369</sup> just as Khan called for rights-based multidimensional and multi-stakeholder responses to tackle online disinformation.

---

<sup>363</sup> Interview 1.

<sup>364</sup> De Schutter, “Towards a New Treaty,” 54.

<sup>365</sup> Benesch, “But Facebook’s,” 89 ; *UNHRC A/HRC/38/35*.

<sup>366</sup> Engadget, “What Facebook.”

<sup>367</sup> Andreassen & Vinh, “Business’ Duties Across Borders,” 10.

<sup>368</sup> Ruggie, “A UN Business and Human Rights Treaty?,” 3.

<sup>369</sup> Deva, *Regulating Corporate*, 201.

## 5 Conclusion

In her report on ‘Disinformation and freedom of opinion and expression’, Irene Khan underlined the threats posed by disinformation to human rights in our current digital world and called for better responses from States, but also from companies. With social media increasingly shaping our daily lives and public spheres, ensuring human rights accountability on these platforms is crucial. As corporations operate the most prevalent platforms, the UNGP framework provides one possible tool to do so.

This research focused on a case study of state-led disinformation campaigns as a free speech infringement involving the corporate responsibility of Facebook (now known as Meta) as a social media business. The case retraces how the Malagasy state and its proxies leverage online attacks on Facebook.com as a way to silence dissent, and the research aimed to assess whether Facebook has responded adequately to the resulting impacts on free speech which it is directly linked to. Our main research question (to what extent is Facebook, Inc. responsible for harms to the right to freedom of expression caused by state-led disinformation campaigns deployed on their Facebook platform in Madagascar?) segues into three sub questions, which we now address once more.

The first sub question concerns the scope and content of Facebook’s responsibilities regarding the right to freedom of expression on its platform. As any other business, Facebook has a responsibility to avoid infringing upon the human rights of others and address adverse human rights impacts with which they are involved. As a social media company, Facebook contends with a “content-based” responsibility to respect freedom of expression, in the sense that the company may find itself involved with free speech infringements resulting from the content published on its platform. The exact contours of this responsibility and the corresponding appropriate actions from the company are difficult to define, as any attempt to address harmful user content may result in countervailing impacts on freedom of expression.

We sought to formalize the scenarios in which the company may be involved in such infringements through a “responsibility scale” adapted from the UNGPs. Applying this scale to our case, we found Facebook directly linked, through its users and its platform, to adverse impacts on free speech such as individual self-censorship and overall pull down on civic space and the public debate. Our analysis presented six cases of coordinated online harassment and disinformation campaigns, which targeted journalists, civil society members, oppo-

sition members and whistleblowers in Madagascar. These cases show the weaponization of the Facebook platform to serve state interests in a way that interferes with individuals' right to free speech.

This observation brought us to our second sub question on measures taken by Facebook to address its involvement in the case. The UNGP provide that a company linked to adverse human rights impacts should seek to mitigate these impacts by exercising leverage where possible and preventing them from recurring. However, we have found that the company has yet to address the cases raised in our study and the resulting adverse free speech impacts in any case-specific way. Our analysis thus drew comparative evidence from the company's previous responses in cases with similar aspects, which show what the company could have done to uphold its responsibility. In these countries, the company did provide efforts, to an extent, to take adequate measures to address its impact. It could have drawn on these measures for possible avenues forward. Beyond actual measures taken, the experience and the knowledge the company has acquired through their HRDD process in these countries should have enlightened it on the risk of its platform being misused to stifle and overwhelm dissenting voices.

With this knowledge, and to answer our last sub question (What additional measures, if any, should the company take in the context of the case?), the company should have applied forward thinking to prevent the impacts identified in these countries from recurring in similar contexts. As such, impacts identified in Myanmar and Sri Lanka for example (where online harm could result in self-censorship of users who would otherwise participate in the public debate) could have been prevented from recurring in Madagascar (where this form of self-censorship has been analyzed as one of the main impacts of the campaigns we studied).

To return to our main question, we found Facebook responsible for harms to the right to freedom of expression caused by state-led disinformation campaigns deployed on their platform in Madagascar. Its platform has been instrumental to facilitate these harms and its insufficient and selective HRDD demonstrate a lacking commitment in making the decisions and investments necessary to ensure respect for human rights. In the case of Madagascar, Facebook did not conduct a HRIA or offer otherwise specific responses, despite a socio-political context marked by vulnerabilities. In addressing disinformation, the company conducts country-specific HRIsAs on a situational rather than systematic basis, despite a recent step in the right direction with the publication of their human rights policy. Facebook has been shown to effect a preferential treatment to countries such as Canada and the United States, while not allocat-

ing the same resources in countries of the Global South. Facebook's negligence is compounded by the company's denial of its responsibility in cases where online disinformation resulted in offline harm. In this regard, Facebook still functions in a way that prioritizes profits over people, which brings us back to the core premise of business and human rights.

This research focused on a specific human rights issue involving the business footprint of Facebook. Throughout this paper, we have endeavored to isolate the relevant aspects of said issue to analyze and derive normative conclusions. However, the human rights risks and misgivings of Facebook run deeper than what could be addressed this paper. The platform's most harmful characteristic remains its algorithmic logic that gives the most reach to the most divisive and polarizing content. However, social media provide new opportunities to leverage rights and bring the powerful to account. For these reasons, it is imperative to search for effective responses to lessen companies such as Facebook's harmful impacts, through multidimensional and multistakeholder responses.

## Table of reference

### Interviews

- Semi-directed Interview n° 1 (Anonymous), conducted on 19 March 2022
- Semi-directed Interview n° 2 (Anonymous), conducted on 24 March 2022
- Semi-directed Interview n° 3 (Anonymous), conducted on 25 March 2022

### Facebook Post data

- Data exhibit 3, Facebook post, published on 02 December 2021, accessed on 25 January 2022.  
[https://www.facebook.com/permalink.php?story\\_fbid=4086410671459247&id=1513800925386914](https://www.facebook.com/permalink.php?story_fbid=4086410671459247&id=1513800925386914)
- Data exhibit 4, Facebook post, published on, accessed on 25 January 2022.  
[https://www.facebook.com/permalink.php?story\\_fbid=3702480963185555&id=1513800925386914](https://www.facebook.com/permalink.php?story_fbid=3702480963185555&id=1513800925386914)
- Data Exhibit 6, Facebook post, published on 30 June 2021, accessed on 25 January 2022.  
[https://www.facebook.com/permalink.php?story\\_fbid=3626464824120503&id=1513800925386914](https://www.facebook.com/permalink.php?story_fbid=3626464824120503&id=1513800925386914)
- Date exhibit 7, Facebook post, published on, accessed on 25 January 2022.  
[https://www.facebook.com/permalink.php?story\\_fbid=3621397574627228&id=1513800925386914](https://www.facebook.com/permalink.php?story_fbid=3621397574627228&id=1513800925386914)
- Data exhibit 8, Facebook post, published on 14 June 2021, accessed on 25 January 2022.  
[https://www.facebook.com/permalink.php?story\\_fbid=3583809485052704&id=1513800925386914](https://www.facebook.com/permalink.php?story_fbid=3583809485052704&id=1513800925386914)
- Data exhibit 10, Facebook post, published on 24 April 2021, accessed on 25 January 2022.  
[https://www.facebook.com/permalink.php?story\\_fbid=3443663372400650&id=1513800925386914](https://www.facebook.com/permalink.php?story_fbid=3443663372400650&id=1513800925386914)
- Data exhibit 14, Facebook post, published on 04 February 2021, accessed on 25 January 2022.  
[https://www.facebook.com/permalink.php?story\\_fbid=3240270119406644&id=1513800925386914](https://www.facebook.com/permalink.php?story_fbid=3240270119406644&id=1513800925386914)
- Data exhibit 16, Facebook post, published on 21 January 2021, accessed on 25 January 2022.  
[https://www.facebook.com/permalink.php?story\\_fbid=3194689630631360&id=1513800925386914](https://www.facebook.com/permalink.php?story_fbid=3194689630631360&id=1513800925386914)
- Data exhibit 52, Facebook post, published on 05 May 2021, accessed on 19 March 2022.  
<https://www.facebook.com/groups/2057683684334116/posts/3549497838486019/>
- Data exhibit 53, Facebook post, published on 04 May 2021, accessed on 19 March 2022. <https://www.facebook.com/groups/616012189011168/posts/832645214014530/>
- Data exhibit 55, Facebook post, published on 11 August 2021, accessed on 19 March 2022.



<https://www.facebook.com/vaovaoandroani/posts/pfbid0fGXfw92uea9Gog7G1upcnN3EsKyVbTEAqiQCZBzxUThTqTDV36Cj2CqEzdqqjTHql>

- Data exhibit 51, Facebook post, published on 01 August 2020, accessed on 19 March 2022.

<https://www.facebook.com/vaovaoandroani/posts/pfbid0fGXfw92uea9Gog7G1upcnN3EsKyVbTEAqiQCZBzxUThTqTDV36Cj2CqEzdqqjTHql>

### Social Media Post references

- Borgia, Gail. "Message à notre cher et beau président..." Facebook post. 21 January 2021. <https://www.facebook.com/gail.borgia/posts/10165458529365506>
- Vaovao Androany. "Transparency International - Hanongam-panjakana no tena iraka nampanaovana azy." Facebook post. 11 August 2021. <https://www.facebook.com/vaovaoandroani/posts/893435571180917>
- Rajoelina, Patrick. "Les dirigeants français invitent de façon massive des opposants aux régimes africains et malgaches." Twitter. 07 October 2021. <https://twitter.com/prajoelina/status/1446160627991470086>
- Transparency International Initiative Madagascar. "La délégation malgache invitée au Sommet France Afrique.." Facebook. 10 October 2021. <https://www.facebook.com/transparencymadagascar/posts/2294816800655757>
- Harry Star. Fake News TVM Facebook video. 29 June 2021. <https://www.facebook.com/ratsira/videos/934784907301933>

### Articles

- Balkin, Jack M. "Digital Speech and Democratic Culture: a Theory of Freedom of Expression for the Information Society". New York University Law Vol. 79, Issue 1 (2004) Yale Law School, Public Law Working Paper No. 63. <http://dx.doi.org/10.2139/ssrn.470842>.
- Bağlayan, Başak. "Searching for Human Rights Norms for Corporate Conduct in Domestic Jurisprudence: A Bottom-Up Approach to International Law." *Nordic Journal of Human Rights* 36, 4 (2018): 371–89. doi:10.1080/18918131.2018.1547533.
- Benesch, Susan. "But Facebook's Not a Country: How to Interpret Human Rights Law for Social Media Companies." *Yale Journal on Regulation Online Bulletin*, vol.3 (2020): 86-111.
- Birchall, David. "Corporate Power over Human Rights: An Analytical Framework." *Business and human rights journal* vol. 6, issue 1 (2021): 42-66.
- Bowen, Glenn A. "Document Analysis as a Qualitative Research Method." *Qualitative research journal*, vol. 9, issue 2 (2009): 27-40.
- Buhmann, Karin. "Analysing OECD National Contact Point Statements for Guidance on Human Rights Due Diligence: Method, Findings and Outlook." *Nordic Journal of Human Rights* 36, 4 (2018): 390–410. doi:10.1080/18918131.2018.1547526.
- Buhmann, Karin and Roxana Olivera. "Corporate Responsibility to Respect Human Rights in regards to Privacy Infringements Involving Photo Posting". *Australian*

*Journal of Human Rights* Vol. 26, Issue 1 (2020): 124 - 141.

- De Schutter, Olivier. "Towards a New Treaty on Business and Human Rights." *Business and Human Rights Journal* Vol. 1, Issue 1 (2016): 41-67. doi:10.1017/bhj.2015.5.
- Deibert, Ronald J. and Rafal Rohozinski. "Liberation vs. Control: The Future of Cyberspace". *Journal of Democracy* Vol. 21, issue 4 (2010): 43–57.
- Douek, Evelyn. "Facebook's 'Oversight Board': Move Fast with Stable Infrastructure and Humility." *North Carolina Journal of Law and Technology* Vol. 21, Issue 1 (2019): 1-78.
- Gajos, Maciej. "Facebook in Myanmar: the challenges and promises of applying the United Nations Guiding Principles on Business and Human Rights to a social media company." *Review of International, European and Comparative Law*, vol. XVIII (2020): 122-148.
- Gong, N., Frank, M., and Mittal, P. "Sybil Belief: A Semi-Supervised Learning Approach for Structure-Based Sybil Detection". *IEEE Transactions on Information Forensics and Security*, Vol. 9 (2014): 976–987.
- Howard, P., Bradshaw S. "The Global Organization of Social Media Disinformation Campaign". *Journal of International Affairs*. volume. 71, issue. 1.5 (2019).
- Helm, Rebecca K. and Nasu, Hitoshi. "Regulatory Responses to 'Fake News' and Freedom of Expression: Normative and Empirical Evaluation." *Human Rights Law Review*, vol. 21 (2021): 308-328.
- Klonick, Kate. "The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression". *Yale Law Journal* Vol. 129 Issue 8 (2020): 2232-2605.
- Lwin, Michaël. "Applying International Human Rights Law for Use by Facebook". *Yale Journal on Regulation* Vol. 38, Issue 53 (2020): 53-85.
- McCorquodale, Robert, Lise Smit, Neely, Stuart and Robin Brooks." Human Rights Due Diligence in Law and Practice: Good Practices and Challenges for Business Enterprises." *Business and human rights journal* vol. 2, issue 2 (2017): 195-224.
- Nuñez, Fernando. "Studies Disinformation Legislation and Freedom of Expression." *UC Irvine Law Review*, vol. 10, issue 2 (2020): 783-798.
- O'Brien Claire Methven and Sumithra Dhanarajan. "The Corporate Responsibility to Respect Human Rights : A Status Review". *Accounting, Auditing and Accountability Journal* Vol. 29, Issue 4 (2016): 542 - 567.
- Ruggie, John, and John F. Sherman. "The Concept of 'Due Diligence' in the UN Guiding Principles on Business and Human Rights: A Reply to Jonathan Bonnitcha and Robert McCorquodale." *European Journal of International Law* Vol. 28, Issue 3 (2017): 921–28. doi:10.1093/ejil/chx047.
- Sander, Barrie. "Freedom of Expression in the Age of Online Platforms: The Promises and Pitfalls of a Human Rights-Based Approach to Content Moderation." *Fordham International Law Journal* vol. 43, 4 (2020): 939-1006.
- Spano, Robert. "Intermediary Liability for Online User Comments under the European Convention on Human Rights". *Human Rights Law Review* Vol. 17, issue 4 (2017): 665-679. DOI:10.1093/hrlr/ngx001.
- Wang, Gang, Manish Mohanlal, Christo Wilson and Xiao Wang. "Social Turing Tests:

Crowdsourcing Sybil Detection". ArXiv:1205.3856 (2012).  
<https://doi.org/10.48550/arXiv.1205.3856>

- Wettstein, Florian. "Human Rights as a Critique of Instrumental CSR: Corporate Responsibility beyond the Business Case". *Notizie di Politeia* Vol. 28, Issue 106 (2012): 18–33.
- Wu, Tim. "Is the First Amendment Obsolete?" *Michigan Law Review*, vol. 117, issue 3 (2018): 547-581.

## Books

- Andreassen, Bård and Vö Khan Vinh. *Duties Across Borders: Advancing Human Rights in Transnational Business*. Intersentia, 2018. Colomina, Carme Héctor Sánchez Margalef, and Youngs, Richard. "The impact of disinformation on democratic processes and human rights in the world". Study. European Parliament's Committee on Foreign Affairs, 2021.
- Copi, Irving, Carl Cohen and Daniel Flage. *Essentials of Logic*. Upper Saddle River, NJ: Routledge Second Edition, 2006.
- Deva, Surya, David Bilchitz. *Human Rights Obligations of Business: Beyond the Corporate Responsibility to Respect?*. New York: Cambridge University Press, 2013.
- Deva, Surya. *Regulating Corporate Human Rights Violations: Humanizing Business*. London. New York: Routledge, 2012.
- Deva, Surya. *Regulating corporate human rights violations: humanizing business*. London: Routledge, 2012.
- Gillespie, Tarleton. *Custodians of the internet: Platforms, content moderation, and the hidden decisions that shape social media*. New Haven, Connecticut: Yale University Press, 2018.
- Greene, Samuel, Gregory Asmolov, Fagan, Adam, Ofer Fridman, and Gjuzelov,. Borjan "Mapping Fake News and Disinformation in the Western Balkans and Identifying Ways to Effectively Counter Them." Study. European Parliament's Committee on Foreign Affairs, 2020.
- Hindman, Matthew and Barash, Vlad. "Disinformation, Fake News and Influence Campaigns on Twitter." Study. Knight Foundation, 2018.
- Klang, Mathias; Andrew Murray. *Human Rights in the Digital Age*. Great Britain: The GlassHouse Press, 2005.
- Morozov, Evgeny. *The Net Delusion: The Dark Side of Internet Freedom*. New York: Public Affairs, 2012.
- Publishing, Walch. *Assessment Strategies for Science: Grades 6-8*. Portland: Walch Publishing, 2004.
- Nygaard, Lynn P. *Writing your master's thesis: from A to Zen*. Los Angeles, California: SAGE, 2017.
- Persily, Nathaniel, and Tucker, Josh A., ed. *Social Media and Democracy: The State of the Field, Prospects for Reform*. Cambridge: Cambridge University Press, 2020.
- Ruggie, John. *Just Business: Multinational Corporations and Human Rights*. New York: W. W. Norton & Company, 2013.

- Wardle, Claire and Derakhshan, Hossein. "Information Disorder: Toward an interdisciplinary framework for research and policy making." Report. Council of Europe, 2017.
- Woolley, Samuel C., Nicholas Monaco and Nyst, Carly. "State-sponsored trolling: How Governments Are Deploying Disinformation as Part of Broader Digital Harassment Campaigns". Report. Institute for the Future, 2018.

### Chapters in books

- Andreassen, Bård A. "Comparative analyses of human rights performance." In *Research Methods in Human Rights: A Handbook*, edited by Bård A Andreassen, Hans-Otto Sano, and Siobhán McInerney-Lankford. 222-252. Cheltenham, Gloucestershire: Edward Elgar Publishing, 2017.
- Andreassen, Bård A. and Vinh Vo Khanh. "Business' Duties Across Borders: The New Human Rights Frontier." In *Duties across borders: advancing human rights in transnational business*, edited by Bård A Andreassen and Vo Khanh Vinh. 1-20. Cambridge: Intersentia, 2016.
- Bennett, W. Lance, and Livingston, Steven. "A Brief History of the Disinformation Age: Information Wars and the Decline of Institutional Authority." In *The Disinformation Age: Politics, Technology, and Disruptive Communication in the United States*, edited by W. Lance Bennett and Steven Livingston. 3-40. Cambridge: Cambridge University Press, 2020.
- Bilchitz, David. "A chasm between 'is' and 'ought'? A critique of the normative foundations of the SRSG's Framework and the Guiding Principles." In *Human Rights Obligations of Business: Beyond the Corporate Responsibility to Respect?* edited by Surya Deva and David Bilchitz. 107-137. New York: Cambridge University Press, 2013.
- Bilchitz, David. "Putting Flesh on the Bone: What Should a Business and Human Rights Treaty Look Like?" In *Building a Treaty on Business and Human Rights: Context and Contours*, edited by Surya Deva and David Bilchitz. 1-24. Cambridge: Cambridge University Press, 2017.
- Joseph Sarah. "The Human Rights Responsibilities of Media and Social Media Businesses." In *Human Rights and Non-State Actors*, edited by Stephanie Farrior. Forthcoming.
- Jørgensen, Rikke Frank. "Human Rights and Private Actors in the Online Domain". In *New Technologies for Human Rights Law and Practice*, edited by Molly K. Land, Jay D. Aronson. 251-53. Pennsylvania: Carnegie Mellon University, 2018.
- Rohn, Ulrike. "Social Media Business Models." In *International Encyclopedia of Digital Communication and Society*, edited by Peng Hwa Ang, Robin Mansell. 1-12. Wiley Blackwell and International Communication Association, 2015.

### Phd/Masters thesis/Study

- Ash, Timothy Garton, Robert Gorwa and Danaë Metaxa. *GLASNOST ! Nine Ways Facebook Can Make Itself a Better Forum for Free Speech and Democracy, An Oxford - Stanford Report*. Reuters Institute for the Study of Journalism and University of Oxford, 2019.
- Bradshaw, Samantha and Philip N. Howard. "Troops, Trolls and Troublemakers: A

Global Inventory of Social Media Manipulation". Computational Propaganda Research Project Briefing Paper no. 2017.12, 2017.

- Bradshaw, Samantha and Philip N. Howard. "Challenging Truth and Trust: A Global Inventory of Organized Social Media Manipulation". Project on Computation Propaganda Working Paper. 2018.1. Oxford UK: Oxford Internet Institute, 2018.
- Caplan, Robyn. "Content or Context Moderation ? Artisanal, Community-reliant, and Industrial Approaches". Report. Data & Society, 14 November 2018.
- Chapoy, Thibault. "Etude sur le paysage médiatique à Madagascar". Report. Internews, 2021.
- Lévesque, Maroussia. "Applying the UN Guiding Principles on Business and Human Rights to Online Content Moderation". (19 February 2021). Available at SSRN. <http://dx.doi.org/10.2139/ssrn3789311>
- Randriantavy, Lovamalala and Iloniaina Alain. "Les impacts de la structure de la propriété des médias, la représentation politique et le travail du journaliste." Report. Friedrich Ebert Stiftung, 2016.
- Samaratunge, Shilpa and Sanjana Hattotuwa. *Liking Violence : A Study of Hate Speech on Facebook in Sri Lanka*. Report. Centre for Policy Alternatives, 24 September 2014.
- Ulrich, George. "Research ethics for human rights researchers." In *Research Methods in Human Rights: A Handbook*, edited by Bård A Andreassen, Hans-Otto Sano, and Siobhán McInerney-Lankford. 192-221. Cheltenham, Gloucestershire: Edward Elgar Publishing, 2017.

### Web pages

- 2020 Ranking Digital Rights Corporate Accountability Index. <https://rankingdigitalrights.org/index2020/companies/Facebook>
- ABC. "Social Media Could Hurt Jill Meagher Case". Last updated 1 October 2012. <https://www.abc.net.au/lateline/social-media-could-hurt-jill-meagher-case/4290126>
- ABS-CBN News. "Rappler Boss Condemns 'Patriotic Trolling' on Social Media." 30 January 2018. <http://news.abs-cbn.com/video/news/01/30/18/rappler-boss-condemnpatriotic-trolling-on-social-media>
- Access Now. "After Nearly Two Decades in the Dark, Facebook Releases its Human
- ADL. "Facebook's Transparency Reporting Continues to Obscure the Scope and Impact of Hate Speech." 20 November 2020. <https://www.adl.org/blog/facebooks-transparency-reporting-continues-to-obscure-the-scope-and-impact-of-hate-speech>
- Africanews. "Madagascar: nouveau code controversé de la communication". Last Updated 07 July 2016. <https://fr.africanews.com/2016/07/07/madagascar-le-nouveau-code-controverse-de-la-communication/>
- Amazon. "Amazon Global Human Rights Principles." 2022. <https://sustainability.aboutamazon.com/people/human-rights/principles>
- Amnesty International. "Amnesty International Report 2021/22: Madagascar 2021." 2022. <https://www.amnesty.org/en/location/africa/southern-africa/madagascar/report-madagascar/>

- Amnesty International. "Madagascar. Un journaliste relaxé près de deux ans après son arrestation". 17 April 2019. <https://www.amnesty.org/fr/wp-content/uploads/sites/8/2021/05/AFR3502182019FRENCH.pdf>
- Amnesty International. "Madagascar: Authorities Must Drop Trumped-up Charges Against Whistleblower Ravo Ramasomanana." 25 November 2021. <https://www.amnesty.org/en/documents/afr35/5045/2021/en/>
- Amnesty International. "Madagascar: Global Leaders Must Act Urgently to Save Lives and Protect Rights Threatened by Climate Crisis." 27 October 2021. <https://www.amnesty.org/fr/latest/news/2021/10/madagascar-global-leaders-must-act-urgently-to-save-lives/>
- Amnesty International. "Madagascar: Urgent Humanitarian Intervention Needed as Millions Face Hunger due to Devastating Famine." 21 May 2021. <https://www.amnesty.org/en/latest/news/2021/05/madagascar-urgent-humanitarian-intervention-needed-as-millions-face-hunger-due-to-devastating-famine-2/>
- AP News. « People of Profit? Facebook papers show deep conflict within.» Last modified on 25 October 2021. <https://apnews.com/article/the-facebook-papers-whistleblower-misinfo-trafficking-64f11ccae637cdfb7a89e049c5095dca>
- Apple. "Our Commitment to Human Rights." 2020. [https://s2.q4cdn.com/470004039/files/doc\\_downloads/gov\\_docs/Apple-Human-Rights-Policy.pdf](https://s2.q4cdn.com/470004039/files/doc_downloads/gov_docs/Apple-Human-Rights-Policy.pdf)
- Article One. "Assessing the Human Rights Impacts of Facebook's Platforms in Indonesia." 2018. <https://about.fb.com/wp-content/uploads/2020/05/Indonesia-HRIA-Executive-Summary-v82.pdf>
- Article One. "Assessing the Human Rights Impacts of the Facebook Platform in Sri Lanka." 2018. <https://about.fb.com/wp-content/uploads/2020/05/Indonesia-HRIA-Executive-Summary-v82.pdf>
- Article One. "Facebook - Challenge: Facebook Inc. Partnered with Article One to Surface and Mitigate Risks Related to its Platforms in Sri Lanka and Indonesia." 2022. <https://www.articleoneadvisors.com/case-study-facebook>
- Association for Progressive Communications. "Release of the Human Rights Impact Assessment of Facebook in India". Last updated 21 January 2022. [https://www.apc.org/en/pubs/release-human-rights-impact-assessment-facebook-in-dia#:~:text=In%202020%2C%20Facebook%20\(now%20Meta,has%20not%20yet%20been%20released](https://www.apc.org/en/pubs/release-human-rights-impact-assessment-facebook-in-dia#:~:text=In%202020%2C%20Facebook%20(now%20Meta,has%20not%20yet%20been%20released)
- Association for Progressive Communications. "Release of the Human Rights Impact Assessment Report of Facebook in India." 21 January 2022. <https://www.apc.org/en/pubs/release-human-rights-impact-assessment-facebook-india>
- BBC News. "Do Not Use Untested Remedies, WHO Africa Warns." 5 May 2020. <https://www.bbc.com/news/world-africa-52544187>
- BBC. "Matter of Fact-checkers: Is Facebook Winning the Fake News War ?" 2 April 2019. <https://www.bbc.com/news/technology-47779782>
- BBC. "Ryan Giggs can be Legally Named as "Affair" Footballer". 21 February 2012. <https://www.bbc.com/news/uk-17114875>

- BBC. "Facebook changes its name to Meta in major rebrand." Last modified on 28 October 2021. <https://www.bbc.com/news/technology-59083601>
- BBC. "Madagascar ex-leader Andry Rajoelina wins presidential run-off vote." Last modified on 27 December 2018. <https://www.bbc.com/news/world-africa-46694430>
- Bloomberg. "Facebook's First Human Rights Chief Confronts Its Past Sins." 29 January 2020. <https://www.bloomberg.com/news/articles/2020-01-28/facebook-s-first-human-rights-chief-seeks-to-tame-digital-hate>
- BSR. "A Human Rights Impact Assessment of Facebook in Cambodia." 12 May 2020. [www.bsr.org/en/our-insights/blog-view/a-human-rights-impact-assessment-of-facebook-in-cambodia](http://www.bsr.org/en/our-insights/blog-view/a-human-rights-impact-assessment-of-facebook-in-cambodia)
- BSR. "A Human Rights Impact Assessment of Facebook in Cambodia." 12 May 2020. [www.bsr.org/en/our-insights/blog-view/a-human-rights-impact-assessment-of-facebook-in-cambodia](http://www.bsr.org/en/our-insights/blog-view/a-human-rights-impact-assessment-of-facebook-in-cambodia)
- BSR. "Human Rights Impact Assessment Executive Summary: Facebook in Myanmar." December 2019. [https://about.fb.com/wp-content/uploads/2018/11/bsr-facebook-myanmar-hria\\_final.pdf](https://about.fb.com/wp-content/uploads/2018/11/bsr-facebook-myanmar-hria_final.pdf)
- BSR. "Human Rights Impact Assessment Executive Summary: Facebook in Cambodia." December 2019. [https://about.fb.com/wp-content/uploads/2020/05/BSR-Facebook-Cambodia-HRIA\\_Executive-Summary2.pdf](https://about.fb.com/wp-content/uploads/2020/05/BSR-Facebook-Cambodia-HRIA_Executive-Summary2.pdf)
- Business and Human Rights Center. "Professor John Ruggie Calls Upon Facebook to Make Significant Changes to Align its Practices with the UNGPS & Prevent it to be Used to Incite Violence". 20 November 2018. <https://www.business-humanrights.org/en/latest-news/professor-john-ruggie-calls-upon-facebook-to-make-significant-changes-to-align-its-practices-with-the-ungps-prevent-it-being-used-to-ignite-violence/>
- China Digital Times. "Youth Volunteers to Spread Sunshine Online". 13 April 2015. <https://chinadigitaltimes.net/2015/04/translation-youth-volunteers-to-spread-sunshineonline/>
- CNN. "Facebook Knew it was being used to incite violence in Ethiopia. It did little to stop the spread, document show." Last Updated 25 October 2021. <https://edition.cnn.com/2021/10/25/business/ethiopia-violence-facebook-papers-cmd-intl/index.html>
- Columbia Journalism Review. "Facebook's Fact-checking Program Falls Short" 2 August 2019. [https://www.cjr.org/the\\_media\\_today/facebook-fact-checking.php](https://www.cjr.org/the_media_today/facebook-fact-checking.php)
- E.Africa Journalists Network. "840 Cases in Madagascar." <https://eajournalistsnetwork.org/madagascar/>
- Engadget. "What Facebook Should Change, according to its Whistleblower." 6 October 2021. [https://www.engadget.com/what-facebook-whistleblower-frances-haugen-said-should-change-143051354.html?guccounter=1&guce\\_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce\\_referrer\\_sig=AQAAAMkuJ1M8rx3FrO3C4\\_5Y\\_drQhJEXHSM9p07NhlI73v\\_D3t7QW3NrDjh6IA9tJSDu5Uu9LTtd0a-qRgKP9LeGVOj3Ma5h-3cHzpNdTgxEWu1mKbN8F6mTEbOhOpTaFVe4VWzfXq1osfqFj4k1Eqmh1riK5Zljbl7cpkH3c5oC\\_Ebw](https://www.engadget.com/what-facebook-whistleblower-frances-haugen-said-should-change-143051354.html?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAMkuJ1M8rx3FrO3C4_5Y_drQhJEXHSM9p07NhlI73v_D3t7QW3NrDjh6IA9tJSDu5Uu9LTtd0a-qRgKP9LeGVOj3Ma5h-3cHzpNdTgxEWu1mKbN8F6mTEbOhOpTaFVe4VWzfXq1osfqFj4k1Eqmh1riK5Zljbl7cpkH3c5oC_Ebw)

- Facebook Corporate Human Rights Policy. 2022. <https://about.fb.com/wp-content/uploads/2021/03/Facebooks-Corporate-Human-Rights-Policy.pdf>
- Facebook Privacy Principles. 2022. <https://www.facebook.com/about/basics/privacy-principles>
- Facebook. "Diversity." 2022. <https://diversity.fb.com/>
- Facebook. "Facebook's Response : Cambodia Human Rights Impact Assessment". 12 May 2020. <https://about.fb.com/wp-content/uploads/2021/03/FB-Response-Cambodia-HRIA.pdf>
- Facebook. "Facebook's Response : Cambodia Human Rights Impact Assessment". 12 May 2020. <https://about.fb.com/wp-content/uploads/2021/03/FB-Response-Cambodia-HRIA.pdf>
- Facebook. "Facebook's Response : Sri Lanka Human Rights Impact Assessment". 12 May 2020. <https://about.fb.com/wp-content/uploads/2021/03/FB-Response-Sri-Lanka-HRIA.pdf>
- Facebook. "Information for Law Enforcement Authorities." 2022. <https://www.facebook.com/safety/groups/law/guidelines/>
- FindLaw. "New Targets: An Overview of Firearms Litigation." Last Updated 26 March 2008. <https://corporate.findlaw.com/litigation-disputes/new-targets-an-overview-of-firearms-litigation.html>
- FindLaw. "Responsibility of Firearms Owners and Dealers for their Second Amendment Right to Bear Arms : A Survey of the Caselaw." Last Updated 17 May 2016. <https://corporate.findlaw.com/litigation-disputes/responsibility-of-firearm-owners-and-dealers-for-their-second.html>
- France24. "Facebook 'troll farms' play outsized role in Madagascar's politics." Last modified 05 October 2021. <https://www.france24.com/en/tv-shows/focus/20211005-facebook-troll-farms-play-outsized-role-in-madagascar-s-politics>
- Freedom House. "Freedom in the World 2021: Madagascar". <https://freedomhouse.org/country/madagascar/freedom-world/2021>
- Global Network Initiative. "Protecting and Advancing Freedom of Expression and Privacy in the ICT Sector." 2022. <https://globalnetworkinitiative.org/>
- Global Network Initiative. "The GNI Principles". May 2017. <https://globalnetworkinitiative.org/gni-principles/>
- Go Banking Rates. "How Much is Facebook Worth?" 8 February 2022. <https://www.gobankingrates.com/money/business/how-much-is-facebook-worth/>
- Google. "Human Rights." 2022. <https://about.google/human-rights/>
- Hattotuwa, Sanja. "Making Facebook's New Human Rights Policy Real" Commentary. Institute for Human Rights and Business, 20 April 2021. <https://www.ihrb.org/focus-areas/information-communication-technology/commentary-making-facebook-new-human-rights-policy-real>
- Haugen, Frances. "Is Facebook Undermining Democracy and Creating Harm?" 01 April 2022. <https://www.youtube.com/watch?v=CcC3VPqQqBw>
- Higg Brand and Retail Guidance. "Understanding Human Rights Due Diligence." 2019. <https://howtohigg.org/wp-content/uploads/2019/10/1.-Understanding-Human-Rights-Due-Diligence-Final.pdf>



- <https://www.nobelprize.org/prizes/peace/2021/press-release/>
- Institut Français des Relations Internationales. "Nouveau Sommet Afrique-France: la continuité masquée de la politique africaine d'Emmanuel Macron." 27 October 2021. <https://www.ifri.org/fr/publications/editoriaux-de-lifri/lafrique-questions/nouveau-sommet-afrique-france-continue-masquee>
- Institute for Human Rights and Business. "Making Facebook's Human Rights Policy Real." 20 April 2021. <https://www.ihrb.org/focus-areas/information-communication-technology/commentary-making-facebook-new-human-rights-policy-real>
- Internet Freedom Foundation. "Examining Facebook's Human Rights Policy." 30 March 2021. <https://internetfreedom.in/facebook-human-rights-policy/>
- Jeune Afrique. "Madagascar: levée des saisies des biens de Mamy Ravatomanga en France." 15 July 2019. <https://www.jeuneafrique.com/803651/societe/madagascar-levee-des-saisies-des-biens-de-mamy-ravatomanga-en-france/>
- Jeune Afrique. "Rasoanaivo: à Madagascar, l'article 20 de la loi sur la cybercriminalité légalise la dictature." 5 August 2014. <https://www.jeuneafrique.com/47575/politique/rasoanaivo-madagascar-l-article-20-de-la-loi-contre-la-cybercriminalite-l-galise-la-dictature/>
- Meta. "Our Commitment to Human Rights". Last modified on 16 March 2021. <https://about.fb.com/news/2021/03/our-commitment-to-human-rights/>
- Jurist. "Civil Society Questions Facebook for Delaying India Human Rights Impact Assessment Report." 22 January 2022. <https://www.jurist.org/news/2022/01/civil-society-questions-facebook-for-delaying-india-human-rights-impact-assessment-report/>
- L'Express de Madagascar. "Injection contre le coronavirus - coup d'envoi des essais cliniques." 27 May 2020. <https://lexpress.mg/27/05/2020/injection-contre-le-coronavirus-coup-denvoi-des-essais-cliniques/>
- L'Express de Madagascar. "Procès - Maître Francesca écope d'une peine de sursis." 23 February 2021. <https://lexpress.mg/23/02/2021/proces-maitre-francesca-ecope-dune-peine-de-sursis/>
- Meta AI. "Collaborating on the Future of AI Governance in the EU and around the World." 15 June 2020. <https://ai.facebook.com/blog/collaborating-on-the-future-of-ai-governance-in-the-eu-and-around-the-world/>
- Meta. "An Independent Assessment of the Human Rights Impact of Facebook in Myanmar." 5 November 2018. <https://about.fb.com/news/2018/11/myanmar-hria/>
- Meta. "Facebook Community Standards." 2022. <https://transparency.fb.com/en-gb/policies/community-standards/?source=https%3A%2F%2Fwww.facebook.com%2Fcommunitystandards%2F>
- Meta. "How Facebook's Third Party Fact-checking Works?" 1 June 2021. <https://www.facebook.com/journalismproject/programs/third-party-fact-checking/how-it-works>
- Meta. "How We Respond to Unauthentic Behaviour on our Platforms: Policy Update" 21 October 2019. <https://about.fb.com/news/2019/10/inauthentic-behavior-policy-update/>

- Meta. "Meta's Third Party Fact-checking Program" 2022. <https://www.facebook.com/journalismproject/programs/third-party-fact-checking>
- Meta. "Pays qui utilisent la vérification des information - Meta Journalism Project" 2022. [https://www.facebook.com/journalismproject/introducing-facebook-journalism-project?locale=fr\\_FR](https://www.facebook.com/journalismproject/introducing-facebook-journalism-project?locale=fr_FR)
- Meta. "Responsible Supply Chain." 2022. <https://sustainability.fb.com/responsible-supply-chain/>
- Meta. "Transparency Reports." 2022. <https://transparency.fb.com/data/?source=https%3A%2F%2Ftransparency.facebook.com%2F>
- Microsoft. "Microsoft Global Human Rights Statement." 2022. [https://www.microsoft.com/en-us/corporate-responsibility/human-rights-statement?activetab=pivot\\_1:primaryr7](https://www.microsoft.com/en-us/corporate-responsibility/human-rights-statement?activetab=pivot_1:primaryr7)
- Midimadagasikara. "Code de la communication: La loi liberticide validée par la HCC". 13 August 2016. <https://midi-madagasikara.mg/2016/08/13/code-de-communication-loi-liberticide-validee-hcc/>
- National Action Plans on Business and Human Rights. "UN Guiding Principles: Principle 11 - Commentary". <https://globalnaps.org/ungp/guiding-principle-11/>
- Open Democracy. "In the Crosshairs of Azerbaijan's Patriotic Trolls." 22 November 2016. <https://www.opendemocracy.net/odrussia/arzu-geybulla/azerbaijan-patriotic-troll>
- OpinioJuris. "Facebook's Commitment to Human Rights." 16 March 2021. <http://opiniojuris.org/2021/03/16/facebooks-commitment-to-human-rights/>
- Politico. "The Long and Brutal History of Fake News." Published 18 December 2016. <https://www.politico.com/magazine/story/2016/12/fake-news-history-long-violent-214535/>
- Politikà. "Etude sur le monde des médias à Madagascar: un paysage médiatique concentré et déséquilibré." 23 August 2021. <https://politika.mg/17351/>
- Quartz. "Facebook Enabled Human Rights Abuses in Myanmar, but Promises to do Better Next Time." 6 November 2018. <https://qz.com/1452551/facebook-enabled-human-rights-abuses-in-myanmar-promises-reform/>
- Reporters Sans Frontières. "Madagascar: des médias sous influence". <https://rsf.org/fr/pays/madagascar>
- Reporters Sans Frontières. "RSF dénonce une loi criminalisant les délits d'expression sur Internet." Last Updated 20 January 2016. <https://rsf.org/fr/rsf-denonce-une-loi-criminalisant-les-delits-dexpression-sur-internet>
- Reporters Sans Frontières. "Smear Campaign Targets Prize-winning Reporter." 1 July 2021. <https://rsf.org/en/smear-campaign-targets-prize-winning-reporter-madagascar>
- Rest of World. "The Facebook Papers reveal staggering failures in the Global South." 26 October 2021. <https://restofworld.org/2021/facebook-papers-reveal-staggering-failures-in-global-south/>
- Reuters. "Apple becomes first company to hit \$3 trillion market value then slips." 4 January 2022. <https://www.reuters.com/markets/europe/apple-gets-closer-3-trillion-market-value-2022-01-03/>

- Reuters. "Facebook Says it Suspends Accounts tied to Putin Ally for Meddling in Africa." 30 October 2019. <https://www.reuters.com/article/facebook-accounts-russia-idINL8N27E68N>
- RFI. "Madagascar: Verdict dans l'affaire du lanceur d'alerte du ministère de la Santé." 30 November 2021. <https://www.rfi.fr/fr/afrique/20211130-madagascar-verdict-dans-l-affaire-du-lanceur-d-alerte-du-minist%C3%A8re-de-la-sant%C3%A9>
- RFI. "Présidentielle à Madagascar: faux comptes et rumeurs prolifèrent sur Facebook." 13 December 2018. <https://www.rfi.fr/fr/afrique/20181213-presidentielle-madagascar-faux-comptes-rumeurs-prolifere-facebook>
- Rights Policy. "16 March 2021. <https://www.accessnow.org/facebook-human-rights-policy/>
- Science. "Unproven Herbal Remedy Against Covid-19 Could Fuel Drug Resistant Malaria, Scientists Warn." 6 May 2020. <https://www.science.org/content/article/unproven-herbal-remedy-against-covid-19-could-fuel-drug-resistant-malaria-scientists#:~:text=Science%20's%20COVID%2D19%20reporting,has%20yet%20to%20be%20shown>
- Statista. "Number of monthly active Facebook Users Worldwide as of 1st Quarter 2022". 28 April 2022. <https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-world-wide/#:~:text=How%20many%20users%20does%20Facebook,used%20online%20social%20network%20worldwide>
- The Guardian. "Facebook is 'biased against facts', says Nobel Prize Winner." 9 October 2021. <https://www.theguardian.com/technology/2021/oct/09/facebook-biased-against-facts-nobel-prize-winner-philippines-maria-ressa-misinformation>
- The Guardian. "Facebook may 'Block Content' Claim as Speculation Grows over Entry into China". 20 April 2011. <https://www.theguardian.com/technology/2011/apr/20/facebook-considers-censorship-claim-china>
- The Guardian. "Madagascar Hands Out 'Miracle' Virus Cure as it Lifts Lockdown." 23 April 2020. <https://guardian.ng/news/madagascar-hands-out-miracle-virus-cure-as-it-lifts-lockdown/>
- The Guardian. "They don't care': Facebook Factchecking in Disarray as Journalists Push to Cut Ties" 13 December 2018. <https://www.theguardian.com/technology/2018/dec/13/they-dont-care-facebook-fact-checking-in-disarray-as-journalists-push-to-cut-ties>
- The Irish Times. "WHO Sceptical as Coronavirus 'Cure' Distributed in Africa." 11 May 2020. <https://www.irishtimes.com/news/world/africa/who-sceptical-as-coronavirus-cure-distributed-in-africa-1.4250671>
- The New York Times. "Where Countries are Tinderboxes and Facebook is a Match." 21 April 2018. <https://www.nytimes.com/2018/04/21/world/asia/facebook-sri-lanka-riots.html>

- The Nobel Prize. "The Nobel Peace Prize 2021." Last modified on 08 October 2021.
- The Santa Clara Principles on Transparency and Accountability in Content Moderation. "An Open Letter to Mark Zuckerberg." 2022. <https://santaclaraprinciples.org/open-letter/>
- The Santa Clara Principles on Transparency and Accountability in Content Moderation, 2018. <https://santaclaraprinciples.org/>
- The Verge. "Facebook Oversight Board Overturns Hate Speech and Pandemic Misinformation Takedown." 28 January 2021. <https://www.theverge.com/2021/1/28/22254155/facebook-oversight-board-first-rulings-coronavirus-misinformation-hate-speech>
- The Verge. "Facebook's new Oversight Board is a Wild New Experiment in Platform Governance" 23 October 2020. <https://www.theverge.com/2020/10/23/21530524/facebooks-new-oversight-board-platform-governance>
- The Wall Street Journal. "Facebook Promised Poor Countries Free Internet. People Got Charged Anyway." 24 January 2022. <https://www.wsj.com/articles/facebook-free-india-data-charges-11643035284>
- Time. "Facebook's Ties to India's Ruling Party Complicate its Fight Against Hate Speech." 27 August 2020. <https://time.com/5883993/india-facebook-hate-speech-bjp/>  
The Wall Street Journal. "Facebook is Stifling Independent Report on its Impact in India, Human Rights Group Say." 12 November 2021. <https://www.wsj.com/articles/facebook-is-stifling-independent-report-on-its-impact-in-india-human-rights-groups-say-11636725601>
- Time. "Facebook's Ties to India's Ruling Party Complicate its Fight Against Hate Speech." 27 August 2020. <https://time.com/5883993/india-facebook-hate-speech-bjp/>
- Time. "UN Fact-finders say Facebook Played a 'Determining' Role in Violence against Rohingya". 13 March 2018. <http://time.com/5197039/un-facebook-myanmar-rohingya-violence/>
- Twitter. Hypervisible Farewell Tour status. 25 October 2020. <https://twitter.com/hypervisible/status/1320444710599708677>
- Financial Times. "The Economics of Big Tech." 2022. <https://www.ft.com/economics-of-big-tech>
- UN News. "Madagascar: 'World Cannot Look Away' as 1.3 Million Face Severe Hunger." 18 November 2021. <https://news.un.org/en/story/2021/11/1106132>
- United Nations Human Rights, Office of the High Commissioner. "Myanmar: An UN Fact-finding Mission Releases its Full Account of Massive Violations by Military in Rakhine, Kachin and Shan States." 18 September 2018. <https://www.ohchr.org/en/press-releases/2018/09/myanmar-un-fact-finding-mission-releases-its-full-account-massive-violations?LangID=E&NewsID=23575>
- Ventures. "Madagascar Launches 'Covid-Organics' as a Remedy for the Novel Coronavirus." 22 April 2020. <https://venturesafrica.com/madagascar-launches-covid-organics-as-a-remedy-for-the-novel-coronavirus/>

- Visual Capitalist. "The World's Tech Giants, Compared to the Size of Economies." 7 July 2021. <https://www.visualcapitalist.com/the-tech-giants-worth-compared-economies-countries/>
- Washington Post. "Facebook Whistleblower Frances Haugen's Senate Testimony". Last updated 4 October 2021. <https://www.washingtonpost.com/context/facebook-whistleblower-frances-haugen-s-senate-testimony/8d324185-d725-4d99-9160-9ce9e13f58a3/>

### Treaties

- International Covenant on Civil and Political Rights (ICCPR), opened for signature on 16 December 1966, 999 UNTS 171 (entered into force 23 March 1976).

### Case Law

- European Court of Human Rights Grand Chamber, Case of Delfi AS vs. Estonia, Judgement, No. 64569/09 (16 June 2015).
- Supreme Court of Florida. Case of Kitchen vs. K-Mart Corporation. Decision ruling, 697 So. 2d 1200 (1997) (17 July 1997).
- International Covenant on Civil and Political Rights (ICCPR), opened for signature on 16 December 1966, 999 UNTS 171 (entered into force 23 March 1976).

### UN Documents

- Facebook. *Facebook submission to UN Special Rapporteur on Freedom of Opinion and Expression for Report on Disinformation, annexed to Disinformation and freedom of opinion and expression: Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Irene Khan*. U.N. Doc A/HRC/47/25.
- Human Rights Committee. *General comment No. 34 - Article 19: Freedoms of opinion and expression*, U.N. Doc CCPR/C/GC/34 (September 12, 2011).
- Human Rights Council. *The Promotion, Protection and Enjoyment of Human Rights on the Internet*, U.N. A/HRC/RES/26/13 (14 July 2014).
- Human Rights Council. *The Promotion, Protection and Enjoyment of Human Rights on the Internet*, U.N. Doc A/HRC/RES/32/13 (18 July 2016).
- Human Rights Council. *The Promotion, Protection and Enjoyment of Human Rights on the Internet*, U.N. Doc A/HRC/RES/20/8 (16 July 2012).
- Human Rights Council. *Disease Pandemics and the Freedom of Opinion and Expression*, U.N. A/HRC/44/49 (23 April 2020).
- Human Rights Council. *Clarifying the Concepts of "Sphere of Influence" and "Complicity": Report of the Special Representative of the Secretary General on the Issue of Human Rights and Transnational Corporations and Other Business Enterprises*, John Ruggie, U.N. A/HRC/8/16 (15 May 2008).
- Human Rights Council. *Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, U.N. Doc A/HRC/38/35 (6 April 2018).

- Human Rights Council. Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, U.N. A/HRC/RES/16/4 (8 April 2011).
- Human Rights Council. *Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, U.N. A/HRC/RES/16/4 (8 April 2011).
- Human Rights Council, *Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression: note/by the Secretariat*, U.N. A/HRC/38/35 (6 April 2018).
- Office of the United Nations High Commissioner for Human Rights. *Guiding Principles on Business and Human Rights: Implementing the United Nations “Protect, Respect and Remedy” Framework*, U.N. Doc HR/PUB/11/04 (2011).
- United Nations General Assembly (UNGA). *The right to privacy in the digital age*. G.A. Res. 68/167, U.N. Doc A/RES/68/167 (December 18, 2013).
- United Nations General Assembly (UNGA). *Universal Declaration of Human Rights* U.N. Doc UNGA 217 A (III) (December 10, 1948).
- United Nations Human Rights Council (UNHRC). “Report of the Special Representative of the Secretary General on the issue of human rights and transnational corporations and other business enterprises, John Ruggie: *Business and Human Rights: Further Steps towards the Operationalization of the “Protect, Respect and Remedy” Framework*.” A/HRC/14/27 (April 9, 2010) (SRSG, ‘2010 Report’).
- United Nations Human Rights Council (UNHRC). *Disinformation and freedom of opinion and expression: Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Irene Khan*. U.N. Doc A/HRC/47/25 (April 13, 2021).
- United Nations Human Rights Council (UNHRC). *Promotion and protection of the right to freedom of opinion and expression Note by the Secretary-General*. U.N. Doc A/71/373 (September 06, 2016).
- United Nations Human Rights Council (UNHRC). *Report of the Special Representative of the Secretary-General on the issue of human rights and transnational corporations and other business enterprises – “Human rights impact assessments - resolving key methodological questions”*. U.N. Doc A/HRC/4/74 (February 05, 2007).
- United Nations Human Rights Council (UNHRC). *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*. Res. 32/38, U.N. Doc. A/HRC/32/38 (November 5, 2016).
- United Nations Human Rights Council (UNHRC). *Report of the Special Representative of the Secretary-General on the issue of human rights and transnational corporations and other business enterprises: Guiding Principles on Business and Human Rights: Implementing the United Nations ‘Protect, Respect and Remedy’ Framework*. U.N. Doc A/HRC/17/31 (March 21, 2011).
- United Nations Human Rights Council (UNHRC). *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*. U.N. Doc A/HRC/38/35 (April 06, 2018).

- United Nations Human Rights Council (UNHRC). *Report on freedom of expression, states and the private sector in the digital age, Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*. U.N. Doc A/HRC/32/38 (May 11, 2016).
- United Nations Human Rights Council (UNHRC). *The promotion, protection and enjoyment of human rights on the Internet*. Res. 26/13, U.N. Doc. A/HRC/RES/26/13 (June 26, 2014).
- United Nations Human Rights Council (UNHRC). *The promotion, protection and enjoyment of human rights on the Internet*. Res. 32/13, U.N. Doc. A/HRC/ RES/32/13 (July 18, 2016).

### **Other Documents**

- Albin-Lackey, Chris. "A Failed Approach to Corporate Responsibility." Human Rights Watch, 2022.
- *Corporate Human Rights Policy*, Facebook [Meta].
- *Declaration of Principles on Freedom of Expression and Access to Information in Africa*, adopted on 10 November 2019, African Commission on Human and Peoples' Rights.
- *Declaration of Principles: Building the Information Society: a global challenge in the new Millennium*, published on 12 December 2003, World Summit on the Information Society WSIS-03/GENEVA/DOC/4-E.
- Directive 2000/31/EC of the European Parliament and the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the internal market ('Directive on Electronic Commerce', 8 June 2000 (entered into force 17 July 2000)).
- Healey, Jason. "Beyond Attribution: Seeking National Responsibility in Cyberspace." Issues Brief. Atlantic Council, January 2012.
- International Business Leaders' Forum, International Finance Corporation and United Nations Global Compact. "Guide to Human Rights Assessment and Management." 2010.
- Karatas, Arzum and Serap Sahin. "A Review on Social Bot Detection Techniques and Research Directions." 10th International Information Security and Cryptology Conference, Turkey. 21 October 2017.
- Madagascar, Constitution de la Ivème République (11 décembre 2010).
- Madagascar, Loi No. 2016-029 portant code de la communication médiatisée (12 août 2016).
- Madagascar, Loi No. 2020-006 portant modifications de certaines dispositions de la loi no. 2016-029 portant code de la communication médiatisée (1 July 2020)
- Madagascar, Loi no. 2016-031 modifiant et complétant certaines dispositions de la loi no. 2014-006 du 17 juillet 2014 sur la lutte contre la cybercriminalité (15 July 2016)
- Organisation for Economic co-operation and Development. *OECD Guidelines for Multinational Enterprises*. 2011.
- Organisation for Economic Co-operation and Development. *OECD Due Diligence Guidance for Responsible Business Conduct*. 2018.

- Ruggie, John. "A UN Business and Human Rights Treaty ?" Issues Brief. John F. Kennedy School of Government Harvard Kennedy School, 28 January 2014.
- United Nations Human Rights Office of the High Commissioner (OHCHR), the Organization of Security and Co-operation in Europe (OSCE) Representative of Freedom of Media, the Organization of American States (OAS) Special Rapporteur on Freedom of Expression, the African Commission on Human and People's Rights (ACHPR) Special Rapporteur on Freedom of Expression and Access to Information. *Joint Declaration on Freedom of Expression and "Fake News" Disinformation and Propaganda*. FOM.GAL/3/17 (3 March 2017).
- United States Congress. *Communications Decency Act*, 1996.

**Communication to author**

- E-mail received 5 May 2020.