

CIVIC COMPASS

Un programa de CIVIC HOUSE

Regulating online civic participation platforms:

Should policy-makers discriminate between content moderation in digital social impact platforms vs “Big Techs”?

Di Natale, Marco (Civic House)
Galvan, Cecilia (Civic House, UBA)
Lonati, Maricel (Civic House)
Wright, Gaston (Civic House - UTDT)



Abstract

The advent of digital platforms has opened new avenues for civic engagement, but it has also brought forth challenges related to regulating online content. This paper explores the ongoing debate surrounding internet content regulation and its implications for online civic participation platforms such as Change.org, Avaaz and MoveOn. Specifically, it delves into the crucial role these platforms play in empowering citizens to participate in the political process while considering the delicate balance between safeguarding freedom of speech and addressing concerns about harmful content dissemination. Advocates on one side argue fervently for an open and free internet, where individuals can freely express their opinions without fear of censorship or retaliation. Conversely, there is an increasing concern about the proliferation of false information, hate speech, and other harmful content. To address these competing concerns, numerous countries have implemented various regulations aiming to control the dissemination of harmful content on the internet. These regulations range from legal restrictions on certain types of speech to voluntary content moderation policies enforced by social media companies. This paper highlights the ongoing debate surrounding the impact of content regulations on online civic participation platforms. Some argue that these regulations could hinder the free flow of information, thus restricting citizens' ability to engage in the political process. Others contend that such regulations are essential to protect the public from harmful content and ensure the internet's continued growth as a tool for civic engagement.

¹ We deeply appreciate the financial support provided by Mott Foundation and Luminate Foundation, which made this research possible. Their contribution has played a crucial role in the advancement and application of our understanding in this field.

Introduction²

Social media platforms have become an integral part of modern society, with billions of people worldwide using platforms like Facebook, Twitter, and Instagram to connect with others, share information, and engage in online communities. In recent years, social media platforms have also emerged as powerful tools for promoting social change, with activists, non-profit organisations, and other groups using platforms to raise awareness, mobilise supporters, and advocate for policy change.

While social media platforms offer many opportunities for promoting social change, they also present challenges and risks. For example, social media platforms can be used to spread harmful content, such as hate speech and disinformation.

This paper explores the opportunities and challenges associated with using social media platforms to promote social change, focusing on the role of commercial social media platforms such as Twitter, Facebook, and Instagram and similar non-commercial social change platforms such as Change.org, MoveOn and Avaaz.³

The paper initiates by dissecting diverse content policies employed by social impact platforms, juxtaposing them with those of 'big tech' platforms. Subsequently, the focus shifts towards the intricacies and potential pitfalls tied to the regulation of content on social media platforms, both current and prospective. Policymakers are urged to grasp the distinctions between regulating commercial platforms, where content is monetarily driven, and social change platforms, where content predominantly serves the purpose of enhancing civic engagement, petitioning authorities, and upholding civil rights.

Finally, the paper concludes with a discussion of best practices for using social media platforms to promote social change and recommendations for policymakers and other stakeholders on how to maximise the potential of these platforms while minimising the risks of regulation that might shrink digital civic space or freedom of expression.

² We acknowledge the indispensable contributions of Guillermina Mayorca, Agustina Acosta, Stefania Di Modica and Juan Martín Marinangeli from Torcuato Di Tella University, who provided invaluable support as research assistants throughout the course of this study. Their tireless efforts included assistance with data collection, revising the existing literature, and interviews with experts and staffers from different platforms, and contributions to manuscript preparation and revision. Their extensive expertise in the subject matter, combined with their rigorous attention to detail and steadfast commitment to excellence, were instrumental in the successful completion of this project. We extend our sincere appreciation to Guillermina, Agustina, and Stefania for their outstanding work and dedication to this research.

³ The paper will also use the term "social impact platforms or social change platforms" to refer to non-commercial or civic society digital platforms.

Social Change Platforms and “Big Tech” Platforms: Why are they different?

The distinction between commercial social media platforms, usually named “big techs”, and social change platforms like Change.org, Avaaz or MoveOn is undoubtedly important, and decision-makers are unlikely aware of this difference while trying to regulate “platforms”.

Regulating social media platforms is a complex issue that requires consideration of a range of factors, including the size and scope of the platforms, their business models, and the potential harm they pose to users and society. Decision-makers are unlikely to know that commercial social media platforms like Facebook, Twitter, and Instagram are much larger and more complex than social change platforms like Change.org, MoveOn or Avaaz. They may pose different types of harm to users and society. There is also no clear understanding of the threats to freedom of expression and democratic pluralism in the digital age.

Furthermore, while content on commercial platforms is regarded as a valuable asset integral to their revenue model, social change platforms view content as a catalyst for fostering political accountability and encouraging civic participation. This dichotomy engenders a pivotal discourse concerning content regulation and the fundamental essence of these platforms.

Recently, there has been growing concern about social media platform's role in spreading harmful content, such as hate speech and misinformation (Calvo & Aruguete, 2020). This has led to a increased scrutiny of these platforms by regulators, lawmakers, and the public, and a range of proposals have been put forward for how to address these issues (Green, 2018).

While social change platforms like Change.org, Avaaz or MoveOn may not be subject to the same level of scrutiny as commercial social media platforms, decision-makers might be considering how to regulate these platforms to ensure that they do not contribute to spreading harmful content or other adverse social outcomes.

In some respects, platforms like Change.org and Facebook are similar in that they both allow user-generated content and can potentially spread harmful content, including hate speech, misinformation, and harmful or violent content. However, there are also significant differences between the two platforms.

For example, Facebook is a much larger and more complex platform than Change.org, with billions of users and various services and features. Facebook's business model relies heavily on targeted advertising, which can exacerbate some of the harms associated with the platform, such as the spread of disinformation.

In contrast, Change.org is a smaller and more focused platform primarily geared towards promoting social change through petitions and other forms of activism. While Change.org can also spread harmful content, such as petitions that encourage hate speech or discrimination, the scope and impact of these harms are likely to be more limited than on a platform like Facebook. Moreover, the Change.org business model hinges entirely on donations or support directed towards the social causes championed on the platform

Change.org uses various strategies to target petitions effectively to maximise their impact. Some of the ways they approach petition targeting, are *user-generated petitions*. The platform allows users to create petitions on any issue they deem necessary. By empowering users to start petitions about the causes they care about, the platform ensures that a wide range of topics are covered thus enabling grassroots activism. Another way of creating engagement is *social sharing and networking*: users are encouraged to share petitions through their social networks, including platforms like Facebook, Twitter, email and Whatsapp. This social sharing helps petitions reach a broader audience and increases the chances of gaining signatures and support.⁴

Other strategies include *trending and popular petitions*. The organisation features trending and popular petitions prominently on their platform. These petitions are often highlighted based on the number of signatures, recent activity, or media attention. This helps draw attention to petitions with potentially high impact or widespread appeal. *Personalised recommendations* use algorithms to provide recommendations to users. These recommendations are based on the user's previous actions, interests, and the petitions they have signed or shown interest in. By tailoring recommendations, the platform can suggest relevant petitions to users, increasing the likelihood of engagement.⁵

Geographic targeting is another way of increasing engagement. This allows users to specify the target audience for their petitions based on location. This helps petitions focus on specific geographic areas or communities directly affected by the issue. Finally, media and news coverage are critical to gaining significant traction or reaching milestones. Media coverage can amplify the petition's message and generate broader awareness, potentially influencing decision-makers and attracting more signatures.

⁴ Interview with former Change.org staffer (06/12/2023)

⁵ Interview with former Change.org staffer (06/12/2023)

By employing these targeting strategies, Change.org aims to ensure that petitions receive the necessary visibility and support to have an impact. The platform's broad user base and social sharing capabilities play a crucial role in expanding the reach of petitions and mobilising communities around pressing issues.⁶

Other social change platforms use similar revenue models. For example, Avaaz declares a “member-funded model”⁷, where there is no interference of corporate sponsorship or governments in its funds. This, according to Avaaz, contributes to the independence and accountability of the platform. They rely on the donations of its members, naming the donations “*investment with permanent social change returns.*” They also claim to be regularly audited as a requirement of the law, fiscally responsible, and 100% independent. Although they received initial seed grants from partner organisations, they are now relying only on small online donations to follow only “the people's agenda.”⁹

On the other hand, MoveOn has two separate channels for accepting donations. Members can either donate to Civic Action or Political Action. For these two platforms, there is a gift acceptance policy. The financial resources received “must be consistent with MoveOn’s values.” First, they claim only to receive support for projects consistent with MoveOn’s mission. They do not “*accept any donation that would compromise MoveOn’s independence, or give any donor influence over MoveOn’s mission, goals and campaigns.*” The platform also discloses every contributor to MoveOn Civic Action and Political Action that donated US\$5,000 or more annually. They also declare to not accept money from “for-profit corporations or government entities, as well as entities formed outside of the US.”¹⁰ Both MoveOn and Avaaz claim to be deeply concerned about keeping the platform’s transparency and independence to keep the power in “the people” and avoid influence from big corporations or governments.

Given these differences, it may be appropriate to regulate social change platforms like Change.org and Facebook differently depending on their size, scope, and potential harms. Smaller platforms may be subject to less stringent regulation than larger platforms. Platforms that rely heavily on targeted advertising may be subject to additional regulation to address the harms associated with this business model. Are decision-makers aware of

⁶ Once you've signed a petition, the platform allows users to share the content on Whatsapp, Facebook, Twitter and email.

⁷ Avaaz principles (Visited on July 3rd, 2023)

⁸ Why donate to Avaaz (Visited on July 3rd, 2023)

⁹ Avaaz financial health (Visited on July 3rd, 2023)

¹⁰ MoveOn’s gift acceptance policy (Visited on July 3rd, 2023)

the differences? How might digitalisation and the transition of social change organisations into platforms challenge freedom of expression and a healthy civic space while regulating online content?

Academic research has not addressed civic platform content regulation nor the tremendous growth of the digital space in the citizen sector. The academic literature on civic participation platforms can be classified into two perspectives: government-created civic participation platforms and offline organisations' moving into the digital engagement of citizens.

The first perspective focuses on how governments take the initiative to create platforms where people can give their opinion on projects the government itself presents, as seen by Gilman (2017) or Rodríguez Bolívar (2015). Other articles focus on platforms that allow people to discuss government initiatives despite not being owned by governments. The general perspective of the literature that examines civic platforms is an understanding of civic engagement as the *"opportunities for members of the public to participate in the political process, not only by voting but also by examining one's own and others' viewpoints"* (Nelimarkka et al., 2014: p.2). As a consequence of this perspective, content is not examined as it derives from initiatives presented by government representatives.

The second group of articles written about civic platforms focuses on how organisations that exist offline create an online engagement strategy. This literature analyses the content on civic platforms, not from a regulatory perspective but intending to understand what organisations choose to communicate online and how they achieve engagement with citizens using online mechanisms (Skaržauskienė & Monika, 2020).

Regarding content, research, like regulation, has not distinguished between commercial and non-commercial platforms. Most of what has been written addresses political-related content concerns that arose due to the Cambridge Analytica scandal during the 2016 US Presidential Elections (MacKinnon and Pakzad, 2018). Consequently, the focus is placed on social media platforms like Facebook and Twitter.

As we have seen, academic research about social impact platforms only focuses on these platforms' strategies to increase participation. Content regulation investigation is conducted solely in terms of social media platforms, ignoring the impact of these types of regulations on social impact platforms. The literature review has shown that no specific investigation has been done about how content regulation affects social impact platforms, the difference between them and commercial platforms, and how that differentiation matters when it comes to regulating.

Social Change Platforms and the Emergence of Digital Citizens

The emergence of these platforms in the context of digital activism can be attributed to several factors. Firstly, the rise of the internet and social media provided new avenues for communication and organising. These platforms leveraged the power of technology to connect people globally, allowing them to collaborate and take collective action in ways that were not possible before.

Digital activism, often called "online activism" or "e-activism," gained prominence in the late 1990s and early 2000s. This period saw the proliferation of internet usage and the growth of online communities. As a result, activists began harnessing the internet's potential to raise awareness, mobilise support, and advocate for social change (Castells, 2012).¹¹

Change.org, founded in 2007, emerged during this transformative period. It capitalised on the idea that online petitions could effectively mobilise people and amplify their voices. By creating a user-friendly platform for petition campaigns, Change.org enabled individuals to initiate and promote causes they cared about, attracting a broad range of users and issues.¹² Ben Rattray, the CEO of Change, and Mark Dimas, the company's Chief Technology Officer, started the organisation with the platform's primary objective when it was created to connect users with the social causes they are passionate about. According to its definition, Change.org's mission is to empower people worldwide to make the changes they want to see.¹³ They describe themselves as an open platform because when individuals with diverse experiences and different perspectives can participate in the conversation, more changes can occur. It aims to amplify the voices of individuals and help them drive meaningful change by mobilising supporters and creating public awareness through its online petition campaigns. According to the Executive Director of Argentina Change.org, Latin America has 98 million users.¹⁴

Avaaz also shared a similar vision of leveraging technology for global activism. Launched in 2007, by Ricken Patel, Tom Pravda, Tom Perriello, Eli Pariser, David Madden, Jeremy Heimans, and Andrea Woodhouse,¹⁵ with a simple democratic mission: organise citizens of

¹¹ This book by Manuel Castells explores the role of the internet and digital networks in contemporary social movements, including the rise of online activism.

¹² Change.org | <https://www.change.org/>

¹³ Change.org mission statement (Visited on July 3rd, 2023)

¹⁴ Interview with Change.org Director in Argentina Leandro Asensos.

¹⁵ Avaaz Foundation. Avaaz Facts (Visited on July 3rd, 2023)

all nations to close the gap between the world we have and the world most people everywhere want. It brings together a global community of activists and mobilises them for various causes, including human rights, climate change, corruption, and social justice.

Its founders recognised the potential of online networks to connect activists worldwide and drive change at an international level. Avaaz focused on human rights, environmental protection, and social justice issues, providing a platform for people to participate in campaigns, sign petitions, and coordinate actions. The platform does not release the number of users.

MoveOn, founded in 1998, predates Change.org and Avaaz but is crucial in laying the groundwork for digital activism.¹⁷ It originated as an email petition campaign to oppose President Bill Clinton's impeachment, demonstrating the power of online organising and mobilisation. MoveOn evolved into a progressive advocacy group, using digital tools to mobilise supporters and influence political outcomes.

Collectively, these platforms have contributed to the growth of online activism and reshaped the landscape of social change movements. By providing accessible and user-friendly tools for engagement, they have empowered individuals to become active advocates and fostered the formation of global communities around specific causes.

It is worth mentioning that these platforms operate as intermediaries, connecting individuals with established nonprofit organisations or facilitating direct communication with policymakers.¹⁸ They act as catalysts for change, bridging the gap between individuals and institutions and providing a digital infrastructure for collective action.

The emergence of social change platforms like Change.org, Avaaz, and MoveOn was a response to the opportunities presented by digital technologies and the need for new models of activism in the online age. They revolutionised how individuals engage with social issues, democratising the advocacy process, amplifying the voices of millions worldwide, and developing new mechanisms of digital accountability (Earl & Kimport, 2011)¹⁹.

¹⁶ [Avaaz.org](https://secure.avaaz.org/page/en/) | <https://secure.avaaz.org/page/en/>

¹⁷ MoveOn.org. (2021). Ballotpedia. <https://ballotpedia.org/MoveOn.org>

¹⁸ In the case of Change.org the platform has a functionality where decision-makers can respond to petitions while triggering messages to all petitions signers.

¹⁹ In this book, the authors examine the impact of digital technologies on activism and social movements, providing insights into the rise of online activism during the late 1990s and early 2000s.

The Evolution of Content Policies in Social Change Platforms

Change.org, MoveOn, and Avaaz are all online platforms that allow users to create and sign petitions related to various social, political, and environmental issues. While all three platforms share similar goals, there are some differences in their content policies.

Here are some key differences between the content policies of Change.org, MoveOn, and Avaaz based on a preliminary analysis of their terms of use:

1. Change.org:²⁰ has a broad content policy that allows users to create petitions on any issue as long as it does not violate the platform's community guidelines. These guidelines prohibit hate speech, harassment, and calls for violence. Change.org also has strict rules around the use of copyrighted material in petitions.
2. MoveOn:²¹ has a narrower content policy focusing on progressive causes. The platform encourages users to create petitions on democracy, human rights, social justice, and the environment. MoveOn's content policy prohibits petitions that promote hate speech, discrimination, or violence.
3. Avaaz:²² has a strict content policy prohibiting petitions related to partisan political campaigns or issues unrelated to global social, political, or environmental issues. The platform also prohibits petitions that promote hate speech, violence, or discrimination.

Regarding the type of content allowed on each platform, Change.org has the most permissive content policy, while Avaaz has the most restrictive. MoveOn falls somewhere in between, focusing on progressive causes but with a more permissive content policy than Avaaz. We will draw attention to this initial classification in the following paragraphs.

It is worth noting that all three platforms rely on community moderation to enforce their content policies. Users can report petitions that violate the policies, and platform moderators will review the reports and take appropriate action. The differences between the content policies of Change.org, MoveOn, and Avaaz reflect each platform's different

²⁰Change.org terms of use (Visited on July 3rd, 2023)

²¹MoveOn terms of use (Visited on July 3rd, 2023)

²²Avaaz terms of use (Visited on July 3rd, 2023)

missions and goals. While all three aim to promote social change through online petitions, they approach this issue differently.

However, it is essential to include a more detailed analysis of the terms and conditions stated by these platforms. We will look into the specific wording of the terms and conditions published by Change.org, Avaaz and MoveOn at a later point.

The use of Social Media by Social Change Platforms

The present section explores the evolving dynamics of activism and civic participation in the digital era, analysing the hybrid form of transnational activism. In this part of the paper, we will focus on how Avaaz uses traditional social media platforms to engage NGOs and regular citizens. It also highlights the role of digital media as a channel for achieving social support, fostering citizen-led groups, and addressing the challenges of civic engagement in contemporary political processes.

As Kavada (2012) shows in her article "Engagement, bonding, and identity across multiple platforms: Avaaz on Facebook, YouTube, and MySpace," Avaaz embodies a hybrid form of transnational activism. Unlike the earlier generation of activism characterised by clear political goals and centralised networks, Avaaz combines elements of second-generation activism with traditional advocacy organisations. Avaaz engages in people-driven policy and international influence by leveraging the Internet's power and multiple platforms. Each platform, including Avaaz's website, Facebook and YouTube, offers distinct possibilities for engagement, bonding, and identity building. By utilising pre-existing social networks, its "users can easily become the targets of mobilisation efforts when "friends" forward or broadcast messages about specific campaigns." (Kavada, 2012) Therefore, Avaaz mobilises users through their connections and empowers them to contribute directly to the organisation's agenda. Although individual voices are present, the organisation maintains control over its image and connections through each platform's design, rules and regulations. The dominant voice is always that of Avaaz, ensuring a consistent image across all petitions.

Nonetheless, it should not be overlooked that "moderating comments and managing the interaction on different platforms is also a resource-intensive process and impossible to sustain for some organisations." (Kavada, 2012) However, critics express concerns about the superficiality of online actions, the potential dilution of organisational messages, and the prioritisation of individual voices over collective creativity.

In "Driving civic participation through social media", Williamson (2011) examines the evolution of social involvement in decision-making processes in the social media era. Regarding the overall importance of social media as an element of social mobilisation, the author highlights that *the importance of technological innovation lies in how it facilitates new methods of participation and support for social networks and issues-based agendas.* Williamson finds that traditional NGOs during the late 20th century went through professionalisation processes that put them closer to governmental technocratic discussions but left them out of place for an era where engagement is challenging to sustain. As a consequence of NGOs' shift in profile, Williamson identifies a change in the

perspective of how they build their relationship with society. As the author suggests, it should be thus expected that when the distance between NGOs and society increases, people would no longer be part of the discussion as citizens but become consumers of decisions taken by others.²³

The analysis developed by the author is similar to what Liste and Sorensen (2015) write about governments' behaviour in their digital relation with citizens. In a case study of Norwegian government websites, Liste and Sorestein find that governments create a consumer, citizen or client relation with citizens through websites. What is seen by Williamson in the analysis of the shift in NGO-citizen relations coincides with the consumer perspective described by the authors. In this type of relationship, people are informed about things that happened and given information about the activities of governments. In this case, NGOs do but are not involved in the decision-making process.

Due to this process, civic engagement in political processes is increasingly difficult, especially when seeking to encourage it with the traditional strategies used in the 20th century. In this context, Williamson (2011, p.2) argues, following Coleman (1998) and Putnam (2000), that "social capital is often embodied in the key relationships that exist between individuals or organisations across civil society; access to it is, therefore, negotiated via a range of background factors that include socioeconomic status, geographical circumstances, ethnicity, religion, age, gender and sexual orientation."

Since the relationship between NGOs and citizens partially dissolved at the end of the 20th century, digital media became a channel to achieve social support. In this space, citizen-led groups emerged and became a space of resistance. Especially since society has shown to interact with like-minded people online. The viral, rapid, emergent and connected nature of Web 2.0, understood as user-generated driven pages and platforms, is critical to replicate offline social dynamics and connect loose networks of association such as those resulting from the late 20th social disengagement—platforms like Change.org. Avaaz and MoveOn rely heavily on traditional "BIG tech" social media platforms. For example, 60% of Change.org petitions in Argentina in 2022 came from shares in Whatsapp.²⁴

²³ This process of shifting from citizens to users/consumers has been the subject of extensive literature since the 1990s.

²⁴ Interview with Change.org former staffer in Argentina.

Terms and Conditions in Social Change Platforms

The terms and conditions of the three non-commercial platforms have several points in common. However, a difference can be established in the case of Change.org, identifying it as a more permissive platform than Avaaz and MoveOn.

After changes in their content policies in 2021, we can state that while the platforms share similarities in their objectives, they differ in their content policies and the level of permissiveness when it comes to hosting certain types of petitions. For example, Change.org has a broad approach, allowing users to create petitions on various topics. This permissive nature means petitions on social, political, and personal issues are hosted on the platform. As a result, Change.org's content may encompass diverse subjects, from global political campaigns to personal appeals for local or individual causes. The platform prioritises user-generated petitions and has a large user base across different countries and cultures. It also allows petitions with non-social content²⁵.

Avaaz is an advocacy platform focusing primarily on international human rights, environmental protection, and social justice issues. The platform's content policies are likely to be more focused and selective, with a stronger emphasis on issues that align with its core mission, and the platform is more permissive towards petitions that directly address pressing global concerns and campaigns with a wide impact²⁶.

Finally, MoveOn is a progressive political advocacy organisation and platform. It is more explicitly focused on progressive causes and campaigns. The platform content may be centred around progressive politics, social justice, and environmental activism issues. As a result, the platform may be more permissive towards petitions that align with progressive values and objectives but very restrictive with general issues.²⁷

It is essential to note that these platforms' specific content policies and permissiveness can evolve and may differ based on regional considerations and legal frameworks.

²⁵ Change.org Terms of Use (Visited June 16th, 2023)

²⁶ Avaaz Terms of Use (visited June 16th, 2023)

²⁷ MoveON Terms of Use (visited June 16th, 2023)

How do Community Guidelines and Terms of Conditions work on social impact platforms?

Firstly, it is clarified that none of the platforms explicitly state the specific modifications made when updating the community guidelines and the terms and conditions. Only the date of the last update is mentioned on the official page.

Change.org community guidelines, according to their page, were last updated in June 2020. They determine specific actions that should not be done or could lead to the elimination of petitions or access to accounts being limited²⁸. A member must not, in their use of the platform: generate violence; or incite hatred (by publishing content that attacks anyone based on their: age, colour, disability, ethnic origin, gender, nationality, national origin, race, religion, gender, sexual orientation, medical condition, or veteran status); include symbols and logos from recognised hate groups or individuals associated with them; deceive others; carry out identity theft; violate privacy; carry out aggression and intimidation; include unnecessary explicit content; harm to children; spam; share personal information; perform illegal activities.

If the content violates the Community Guidelines or is inappropriate or objectionable, they may remove it from the platform or edit it to ensure compliance with the rules. In most cases, they will notify the member of their violation and try to work with them to resolve the issue before taking any of these actions (however, they reiterate that they are not obligated to do so²⁹).

Regarding content standards, Avaaz has a stricter policy compared to Change.org. It is displayed on the official website as last updated in April 2019. Similar to Change.org, users must not use this platform to deceive or mislead any person; promote any illegal activity, or advocate, promote or assist any unlawful act; impersonate any person; promote sexually explicit or pornographic material, violence, or discrimination based on race, sex, religion, nationality, disability, sexual orientation or age.

²⁸ According to an interview with a former Change.org staffer, this is observed on a case-by-case basis. They claim to create a context that is as objective as possible about whether the text is infringing the community guidelines.

²⁹ According to an interview with a former Change.org staffer, if certain content is violating the terms of use and there is no response from the user or there is a response but it continues to be a violation of the terms of use, it will be removed. However, there has not been a case in which an user has been banned from the platform.

Furthermore, they add other regulations. The contributions made by users on the platform must not contain any material which is false, defamatory, obscene, indecent, abusive, offensive, harassing, violent, hateful, inflammatory, endangers Avaaz's broader mission, or is otherwise objectionable. Moreover, content must not infringe any patent, trademark, trade secret, copyright or other intellectual property rights of any other person; violate the legal rights (including the rights of publicity and privacy) of others or contain any material that could give rise to any civil or criminal liability under applicable laws or regulations or that otherwise may be in conflict with these Terms of Use and our Privacy Policy. Additionally, they include that content must not involve commercial activities or sales, such as contests, sweepstakes and other sales promotions, barter or advertising, or give the impression that they emanate from us or any other person or entity if this is not the case.

Lastly, users must not use this platform in any way that violates any applicable national, regional, federal, state, local, or international law or regulation; to exploit, harm or attempt to exploit or harm minors in any way by exposing them to inappropriate content, asking for personally identifiable information or otherwise; to transmit, or procure the sending of, any advertising or promotional material, including any "junk mail," "chain letter," "spam," or any other similar solicitation; to harvest or collect email addresses or other personally identifiable information of users from this site by electronic or other means for any purpose; to engage in any other conduct that restricts or inhibits anyone's use or enjoyment of this site, or which, as determined by us, may harm Avaaz or users of the site or expose them to liability; or for any commercial purpose.

MoveOn's content regulation can be considered a middle ground between Change.org and Avaaz's content policies. It is displayed on the official website as last reviewed in April 2022³⁰. If there is inappropriate content on the website, it will be removed at the sole discretion of MoveOn.

MoveOn strongly emphasises that the platform is designed solely for creating, disseminating, and signing non-commercial petitions and raising awareness for public policy issues, policies and legislation. Acceptable petitions, comments, images, and text uploaded or added to the site must be consistent with the site's design. They may be removed if they are inconsistent with such design or violate these Terms of Use³¹.

Like Change.org and Avaaz, users must not, in their use of the platform: promote and/or encourage illegal or unlawful activity in any country; advocate hate, violence, abuse, threats, or harm against any group, individual or entity; harass or encourage others to

³⁰ Avaaz terms of use (Visited on July 3rd, 2023)

³¹ MoveOn privacy and terms of services (Visited on July 3rd, 2023)

harass any group, individual or entity; include content that is vulgar, obscene, profane, pornographic, and/or otherwise offensive; disclose third-party confidential, private, or trade secret information.

Like Avaaz, in no event can this platform be used for any commercial or for-profit endeavour or purpose; or to solicit funds for or contributions to any candidate, group, organisation, committee or entity of any kind. Furthermore, content must not be defamatory, libellous or slanderous. Also, content contributed to the platform must not infringe upon or violate any third-party rights, including others' intellectual property rights, privacy rights, or publicity rights of others, including posting or providing links to unauthorised copies of another party's copyrighted work.

In conclusion, Change.org has the most permissive content policy out of the three non-commercial platforms. Its content policy prohibits violence, inciting hatred, including symbols and logos from recognised hate groups or individuals associated with them, deceiving others, identity theft, violation of privacy, aggression and intimidation, including unnecessary explicit content, harm to children, spam, sharing personal information, and performing illegal activities.

Secondly, MoveOn's content policy is very detailed but less restrictive than Avaaz's. It shares specific critical points with Change.org's content policy but is more comprehensive in such descriptions. On the platform, users must not: promote and/or encourage illegal or unlawful activity in any country; advocate hate, violence, abuse, threats, or harm against any group, individual or entity; harass or encourage others to harass any group, individual or entity; include content that is vulgar, obscene, profane, pornographic, and/or otherwise offensive; or disclose third-party confidential, private, or trade secret information. It also adds that content must not be defamatory, libellous or slanderous.

Moreover, MoveOn's content policy is stricter, as this platform cannot be used for commercial or for-profit purposes. It should, therefore, be used for the creation, dissemination, and signing of non-commercial petitions and to raise awareness for issues relating to public policy issues, policies and legislation. Lastly, content must not infringe upon or violate any third-party rights, including intellectual property rights, privacy rights, or publicity rights of others, including posting or providing links to unauthorised copies of another party's copyrighted work.

Avaaz is the most restrictive of the three non-commercial platforms mentioned. Similar to Change.org and MoveOn, this platform must not be used to deceive or mislead any person; promote any illegal activity, or advocate, promote or assist any unlawful act; impersonate any person; promote sexually explicit or pornographic material, violence, or discrimination based on race, sex, religion, nationality, disability, sexual orientation or age. Like MoveOn, content on the platform must not contain any material which is false, defamatory, obscene,

indecent, abusive, offensive, harassing, violent, hateful, inflammatory, endangers Avaaz's broader mission, or is otherwise objectionable; infringe upon or violate any third-party rights; or involve commercial activities or sales, such as contests, sweepstakes and other sales promotions, barter or advertising.

Furthermore, unlike the other companies' content policies, content must not give the impression that it emanates from the company or any other person or entity if this is not the case. Additionally, it includes that users must not use this platform in any way that violates any applicable national, regional, federal, state, local, or international law or regulation; to exploit, harm or attempt to exploit or harm minors in any way by exposing them to inappropriate content, asking for personally identifiable information or otherwise; to transmit, or procure the sending of, any advertising or promotional material, including any "junk mail," "chain letter," "spam," or any other similar solicitation; to harvest or collect email addresses or other personally identifiable information of users from this site by electronic or other means for any purpose; to engage in any other conduct that restricts or inhibits anyone's use or enjoyment of this site, or which, as determined by us, may harm Avaaz or users of the site or expose them to liability.

Social Change Platforms and Content Removal

The process for removing harmful content or content that violates community guidelines can vary depending on the platform. However, the academic literature and practitioners agree that the procedure involves the following steps before takedown:

1. *Identification of Violations:* Platform moderators or automated systems content that may violate the platform's community guidelines. This could include content that is hate speech, incites violence, promotes terrorism, or is sexually explicit.
2. *Review:* The platform's moderators or automated systems then review the identified content to determine whether it violates community guidelines. This includes an internal process where different stakeholders are consulted. Sometimes, a local representative's intervention is required to provide context and a local interpretation of the possible content violation.
3. *Decision:* Based on the review, the platform will decide on whether to remove the content, leave it up with a warning or label, or take other actions.
4. *Notification:* If the content is removed, the user who posted it is notified of the removal and its reason.
5. *Appeal:* Users may be able to appeal the removal of their content if they believe it was removed in error or the platform's decision was unjust.

Platforms may also have different processes in place for handling specific types of violations or for dealing with repeat offenders. This could also prohibit users from continuing to post content on the platforms. Additionally, some platforms may use machine learning algorithms to identify and flag potentially harmful content, which can speed up the review process (Change.org)

Change.org has more than 450 million users worldwide³². Their content policies have evolved since the global launch of the platform in 2007³³. In recent years the platform has increased the number of staffers working on Policy, which led to more specific and comprehensive content policies. Change.org has a Content Policy that is relatively easy to find in the platform under the Policies section. The section outlines the types of content that are not allowed on the platform, such as hate speech, threats, harassment, and

³² According to statistics site <https://expandedramblings.com/index.php/changeorg-facts-statistics/>

³³ There have been substantial changes in content policy since its original creation. Based on an interview with a former Change.org staffer, content policy centered around whether the content broke the terms of use, while now there is a higher emphasis regarding who created the petition. For example, in the case that a white supremacist or neo-nazi group creates a petition, it will be reviewed.

promotion of violence. When users report content they believe violates the policy, Change.org's moderation team will review it and determine whether it should be removed.

According to Change.org, staffers Change.org follow a five steps process to remove harmful content and enforce their community guidelines.

1. User Reports: Users can report content by clicking a "Report" button on the page where the content appears. This is easy to find on a "petition page."
2. Moderation Review: Change.org's moderation team will review the reported content to determine whether it violates the Content Policy. Intervention from local representatives is critical in making decisions regarding content.
3. Removal Decision: Based on the review, the moderation team will decide on whether to remove the content, leave it up with a warning or label, or take other actions.
4. Notification: If the content is removed, the user who posted it is notified of the removal and its reason. Local teams also play a role in notifying and providing information about why the content has been removed.
5. Appeals: Users can appeal the removal of their content if they believe it was removed in error or the moderation team's decision was unjust.

Change.org also uses machine learning algorithms to help identify potentially harmful content and flag it for review by the moderation team. Additionally, they have a team of human moderators reviewing reported content to ensure it complies with their policies. The global team and local staffers engage with global policy teams to address potential content violations. According to staffers, local representatives play a fundamental role in providing a local understanding of the possible violation. This is a challenge for a multilingual platform with fewer resources to monitor content than commercial platforms such as Facebook, Twitter or Instagram.

Avaaz is a global civic organisation focusing on climate change, human rights, and democracy. They have 69 million users globally³⁴ and 24 million in Latin America. Avaaz's strict content policy prohibits harmful or abusive content, including hate speech, threats, and misinformation. They also have guidelines that prohibit content that promotes violence, harassment, or discrimination based on race, gender, sexual orientation, religion, or any other characteristic. This makes Avaaz a less open platform when compared with Change.org.

Avaaz uses a combination of technology and human review to enforce its content policies.

³⁴ According to Avaaz's website: <https://avaaz.org/page/en/about/>

They employ automated tools to scan for harmful content, such as hate speech. They also have a team of moderators who review flagged content to determine whether it violates its policies. Avaaz encourages its members to report harmful content.

When Avaaz identifies harmful content, they swiftly remove it from their platform. Depending on the severity of the violation, Avaaz may issue a warning, suspend an account, or permanently ban a user from the platform. They also work closely with law enforcement agencies to address illegal activity on their platform.

In conclusion, Avaaz has a robust content policy that prioritises the safety and well-being of its users. They use technology and human review to enforce policies and take swift action to remove harmful content from their platform. While both Avaaz and Change.org are online platforms that allow people to create petitions and campaigns on various issues, their content policies differ in a few key ways.

One of the main differences between the two platforms is the types of content they allow. Change.org has a more permissive content policy that allows a broader range of topics and viewpoints, even controversial ones, as long as they do not violate their guidelines on hate speech, harassment, and violence.

On the other hand, Avaaz has a stricter policy prohibiting harmful or abusive content, including hate speech, threats, and misinformation. Also, Avaaz only allows petitions with a clear social change objective, while Change.org has an open approach where any petition, even those without a clear social change impact. This is based on their philosophy of being a truly open platform. This statement challenges exciting discussions about the role of social change platforms and freedom of expression. For a more detailed explanation of these challenges see the work of Leidner and George and the evolution of digital activism (Leidner & George, 2019).

Another difference is in the enforcement of their content policies. Change.org relies primarily on user reports and identifying harmful content, while Avaaz uses automated tools and human review to scan for and remove harmful content.

Additionally, Change.org's approach to eliminating harmful content is more reactive, while Avaaz takes a more proactive approach by scanning for harmful content before it becomes an issue. Also, as an internal policy, Avaaz takes down non-social change petitions.

While both Avaaz and Change.org are committed to promoting social change and empowering people to make a difference, they have different approaches to content

moderation and ensuring the safety of their users. The differences in their content policies may appeal to diverse audiences and serve other purposes, depending on the type of campaign or petition being created. Avaaz caters to activities and more global issues, while Change.org is more local and focuses on regular citizens.

MoveOn is an interesting example of a social change platform with on campaigning on various topics based on the premise of “championing progressive values”.

“Big Tech” content moderation and takedown

Content moderation is crucial in social media platforms, ensuring user-generated content aligns with community guidelines and standards. This part of the paper examines the content moderation policies of three prominent social media platforms: Facebook, Twitter, and Instagram.

Facebook has developed a comprehensive set of Community Standards that define acceptable content on the platform. These standards cover various categories, including hate speech, violence, nudity, and graphic content. For instance, Facebook prohibits hate speech targeting individuals based on race, ethnicity, religion, and sexual orientation. Violations of these policies can result in content takedowns and user account restrictions.

To enforce content moderation, Facebook employs a combination of user reporting, AI algorithms, and human review processes. When users report content, it goes through a review process to determine its compliance with the Community Standards. Facebook also uses artificial intelligence algorithms to detect and remove violating content proactively, although the system is not without errors and relies on human review for accurate decision-making.³⁵

Facebook has a multi-faceted approach to content takedown, employing a combination of user reporting, artificial intelligence (AI) algorithms, and human review processes. Here is an expanded explanation of Facebook's content takedown approach and process:

1. *User Reporting:* Facebook encourages users to report content that violates the platform's Community Standards. Users can report individual posts, comments, photos, videos, and entire profiles or pages. The reporting feature allows users to select the specific violation category that best matches the content, such as hate speech, graphic violence, nudity, or harassment.
2. *AI Detection:* Facebook employs advanced AI algorithms to detect and proactively identify potentially violating content. These algorithms utilise machine learning techniques and pattern recognition to scan and analyse the vast content posted on the platform. The AI systems are trained on large datasets of previously reported and moderated content to improve their accuracy over time.
3. *Review Process:* Reported and flagged content is reviewed to determine its compliance with Facebook's Community Standards. The review process involves

³⁵ A closer look at Facebook efforts on content moderation can be found on the company annual Human Rights Report <https://humanrights.fb.com/>

human moderators who assess the reported content based on the guidelines provided by Facebook. These moderators are trained to apply consistent and fair judgement when evaluating reported content ^{36 37} .

4. *Escalation and Appeals*: In cases of ambiguity or disagreement about whether reported content violates Community Standards, Facebook provides an escalation and appeals process. Higher-level moderators can review the reported content or undergo a secondary review to ensure fair decision-making. Additionally, Facebook offers an appeals process where users can contest content takedowns or account restrictions if they believe an error has occurred.
5. *Continuous Improvement*: Feedback from users and insights gained from the review process are used to refine the AI algorithms and enhance the efficiency and accuracy of content moderation. Facebook also invests in training its human moderators to ensure consistent enforcement of Community Standards.

It is important to note that content moderation on a platform as vast as Facebook presents significant challenges. The high volume of user-generated content and the need for timely review can sometimes result in mistakes or delays in content takedowns.

Twitter's content moderation policies are outlined in its Rules, which address various issues such as harassment, hate speech, misinformation, and abusive behaviour³⁸. The platform is robust against targeted harassment and abuse, imposing strict penalties on users who engage in such behaviour. Misinformation, particularly regarding public health and election integrity, is another area where Twitter actively engages in content moderation³⁹.

Users can report tweets and accounts for violating Twitter's rules. Twitter employs automated systems and human moderators to review reported content and take appropriate actions, including content takedowns and user suspensions. However, Twitter has faced criticism for inconsistencies in content moderation, with some users alleging biased enforcement of policies based on political or ideological affiliations.

Twitter provides users with a reporting mechanism to flag content violating its Rules. Users can report individual tweets, profiles, direct messages, or other forms of content. Twitter

³⁶ Meta, "How we review content". Link: <https://about.fb.com/news/2020/08/how-we-review-content/>

³⁷ Facebook's Oversight Board. Link: <https://about.fb.com/news/2020/05/welcoming-the-oversight-board/>

³⁸ Platform Use Guidelines, Twitter. (Visited on July 3rd, 2023)

³⁹ Policy regarding misinformation, Twitter. (Visited on July 3rd, 2023)

allows users to specify the violation category, such as harassment, hate speech, violent threats, or misinformation⁴⁰.

Also, the company employs AI algorithms to assist in content moderation. These algorithms use machine learning techniques to analyse and identify potentially violating content. The algorithms consider various signals, including keywords, context, and user behaviour patterns, to detect and prioritise content that may require human review⁴¹.

The use of manual reporting is also essential for Twitter. Reported content undergoes human review by trained moderators. These moderators make decisions regarding content takedowns, warnings, or account suspensions.

Regarding the appeals process for content which has been taken down, Twitter provides an appeals process for users who believe their content was wrongfully flagged or their account was suspended. Users can submit an appeal through the platform's designated channels, and a different team of moderators will review the content or account to ensure fair reconsideration.

Instagram maintains its content moderation policies through its Community Guidelines. It emphasises nudity, self-harm, bullying, and hate speech. The platform prohibits explicit nudity but allows artistic and non-sexual nudity.

Content moderation on Instagram involves a combination of AI algorithms, user reports, and human moderation. The platform utilises AI technology to detect and remove potentially violating content proactively. Users can report content that they believe violates the guidelines, and human moderators review these reports to make final decisions on content takedowns and account actions⁴².

Instagram offers reporting tools for users to flag content that violates its Community Guidelines. Users can report individual posts, comments, stories, profiles, or direct messages. Instagram provides reporting options for violations such as nudity, bullying, hate speech, self-harm, or graphic violence⁴³.

Since the platform is part of META, Instagram utilises AI technology to assist in content moderation. AI algorithms automatically detect and remove violating content, including explicit or sensitive imagery.

⁴⁰ Report breaches, Twitter. (Visited on July 3rd, 2023)

⁴¹ Reporting functions, Twitter. (Visited on July 3rd, 2023)

⁴² How does Instagram Uses Artificial Intelligence to moderate content? Instagram (Visited on July 3rd, 2023).

⁴³ Abuse and spam. Instagram (Visited on July 3rd, 2023).

The algorithms are trained to identify patterns and context indicative of policy violations, aiding in the efficient review and removal of content. Also, Instagram has a team of content reviewers who evaluate reported content and make decisions regarding takedowns, warnings, or account actions. Human moderation helps ensure nuanced judgement in complex cases requiring context-specific analysis⁴⁴.

Content moderation poses significant challenges for these platforms. Striking the right balance between freedom of expression and limiting harmful content remains a constant challenge. The vast amount of user-generated content makes identifying and removing violations difficult.

Additionally, platforms face criticism regarding perceived biases in their content moderation practices, often accused of being too restrictive or not doing enough to combat harmful content.

The debate over the role and responsibility of social media platforms in content moderation continues. Platforms must navigate the fine line between safeguarding user safety and maintaining an open environment for diverse perspectives and discussions. Developing transparent and consistent content moderation policies while addressing concerns of bias and accountability is a crucial challenge.

Facebook, Twitter, and Instagram have implemented comprehensive content moderation policies to foster safe and inclusive online environments, as stated in their Community Guidelines or Terms of Use. These platforms employ a combination of user reporting, AI algorithms, and human moderation to enforce their guidelines. However, challenges persist in striking the right balance and addressing concerns of bias and consistency.

As content moderation evolves, social media platforms will face new challenges. Collaborative efforts involving technology, user feedback, and ongoing policy improvements will be crucial in adapting to the ever-changing landscape (Gillespie, 2018).

While commercial platforms like Facebook, Twitter, and Instagram have well-defined content moderation policies, social change platforms such as Change.org, Avaaz, and MoveOn approach content moderation differently. Unlike commercial platforms, which primarily focus on enforcing community standards to maintain a safe and inclusive environment, social change platforms often prioritise amplifying diverse voices and promoting activism. Content moderation on social change platforms ensures campaigns and petitions align with the platform's mission and values rather than policing user-generated content. These platforms aim to foster a space where individuals can

⁴⁴ Content regulation, Instagram (Visited on July 3rd, 2023).

freely express their opinions and mobilise for social causes. This distinction reflects the ongoing debate surrounding content moderation, highlighting the different priorities and approaches between commercial platforms and platforms dedicated to social change.

However, there is a convergence point between the content regulation mechanisms employed by social change platforms and commercial platforms. For instance, Meta (formerly known as Facebook) partners with civil society organisations worldwide to inform their policy development process. They have established a "Trusted Partner Program" aimed at collaborating with local organisations to enhance their capacity for monitoring social media. This program encompasses 400 non-governmental organisations, including humanitarian and human rights agencies, spanning 113 countries⁴⁵. The primary objective is to gain deeper insights into problematic content trends, prioritise online safety and security, and promote transparent policies. Meta emphasises that these Trusted Partners play a dual role, reporting content and providing crucial feedback on the company's content policies. By incorporating external regulators with unbiased perspectives, commercial companies seek to ensure their content guidelines are not influenced solely by monetary considerations.

⁴⁵Meta Transparency Center <https://transparency.fb.com/>

Regulating Content: The challenges of focusing only on commercial platforms

As discussed in this article, the distinction between civic and commercial platforms is extremely blurry and not addressed by policymakers because of the lack of understanding of the massive transition of civic participation to the digital space.

During the early era of the internet, laws were primarily focused on protecting people's data while navigating the web. However, regulators are now increasingly directing their attention towards regulating the content circulating on the internet. Although only a few countries have passed laws to control platform content, this article proposes a discussion of several laws and projects seeking to do so. Specifically, the following will be discussed: Senator Girardi's project from Chile⁴⁶, Senator Rodriguez Saa's project from Argentina⁴⁷, the "Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet" project from Brazil⁴⁸, the "Digital Services Act" regulation from the EU⁴⁹, and Germany's "Network Enforcement Act".⁵⁰

Although each project or law has slightly different definitions of platforms, they all share a common understanding of platforms as service providers that enable users to share information with others through the internet. This definition applies to any internet service host allowing users to publish and share content with others and view the content shared by others.

The Brazilian project stands out as it seeks to regulate social networking and private messaging services, including encrypted messaging, such as WhatsApp.⁵¹ What the project pretends to regulate in private messaging services is the number of times information is shared, and therefore requires messaging services to develop features that limit the number of times a message can be forwarded. The project also gives messaging

⁴⁶ Senator Girardi's project was presented on September 1st, 2021 and can be downloaded from: <https://www.camara.cl/legislacion/ProyectosDeLey/tramitacion.aspx?prmID=15047&prmBOLETIN=14561-19>

⁴⁷ Senator Rodriguez Saa's project was presented on May 6th, 2020 and can be downloaded from: <https://www.senado.gob.ar/parlamentario/comisiones/verExp/848.20/S/PL>

⁴⁸ Brazil's project was presented on June 20th, 2020. It can be downloaded from: https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra?codteor=1909983

⁴⁹ The Digital Services Act was approved on April 22nd, 2022 and can be downloaded from: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022R2065&qid=1666857835014>

⁵⁰ The Network Enforcement Act was approved in June 2017. The original text can be downloaded from: https://www.bmj.de/SharedDocs/Gesetzgebungsverfahren/Dokumente/NetzDG_engl.pdf;jsessionid=0906A0DC0C2E9320001E64096948A257.2_cid297?__blob=publicationFile&v=2

⁵¹ Social networks and private messaging services are defined in Art. 5°.

services the responsibility of notifying the controlling agency every time a user shares content that is considered illegal in terms of this specific law; something encrypted messaging companies cannot do since they do not access the content of messages shared between users. One last requirement the project stipulates for messaging services is deactivating the default configuration for group chat creation, replacing it with a feature that asks users for consent before letting other users include them in group chats. For private messaging companies and social media platforms, creating a specific team dedicated to content revision and illegal content notification is required.

One significant difference among the projects is how they limit the platforms covered by the regulations. Some projects impose a user threshold, while others apply to all platforms operating within the territory. Germany (2 million)⁵², Brazil (2 million)⁵³, and Argentina (1 million)⁵⁴ limit the application based on the number of users. In contrast, projects from Chile⁵⁵ and regulations in the EU⁵⁶ do not differentiate based on the number of users and apply to every platform.

However, what becomes apparent from a broader perspective is that none of the projects distinguishes between civic and commercial platforms. Platforms such as Facebook, Twitter, Change.org, and Avaaz are treated as equivalent entities for these regulations.

Consequently, the regulations impose the same obligations on all of them. It is worth noting that the case of Chile is exceptional, as the regulation intends to apply to every platform with activity in the country.

Interestingly, while regulators present the project as a response to the need for regulating content on social networking platforms that are incapable of self-regulation, these companies are exempted from any responsibility regarding the content users upload to their servers. The platforms must be neutral regarding content handling when identifying illegal content rather than taking down content considered illegal. Unlike other projects, this one does not mandate platforms to have dedicated content review teams or specify which public office or authority should be responsible for enforcing the law.⁵⁷

Argentina has a project that offers better definitions in terms of scope and objectives but raises concerns regarding freedom of expression. The regulation requires platforms to

⁵² *Network Enforcement Act. Section 1, paragraph 2.*

⁵³ *Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet. Art. 1º.*

⁵⁴ *Rodriguez Saa's project, Argentina. Art. 3º.*

⁵⁵ *Girardi's project, Chile. Art. 2º.*

⁵⁶ *Digital Services Act. Chapter 1, Art. 1º.*

⁵⁷ *Girardi's project, Chile. Art. 6º.*

incorporate a visible tool allowing users to report illegal content. The law defines illegal content as follows: 1) Fake news lacking veracity and transmitted through various communication portals or social networks with the primary intention of intentionally dishonouring or discrediting a person, falsely accusing a person of a specific and circumstantial crime, or inciting panic, anguish, or promoting incorrect behaviour during a pandemic, epidemic, or disease outbreak; 2) Hate speech, which includes expressions or messages intended to intimidate, discriminate, or incite hatred and violence against individuals or groups based on their race, religion, nationality, gender, sexual orientation, disability, or another characteristic.⁵⁸

The decision regarding the illegality of reported content is delegated to an expert team that platforms must establish. This team evaluates whether the reported posts violate content regulations, files the reports in an archive, and protects information related to the reports.⁵⁹

⁵⁸ *Rodriguez Saa's project, Argentina. Art. 1°.*

⁵⁹ *Rodriguez Saa's project, Argentina. Art. 5°.*

Conclusion

In conclusion, social media platforms have become powerful tools for promoting social change, allowing individuals and organisations to raise awareness, mobilise supporters, and advocate for policy change. However, using social media for social change also presents challenges and risks, such as the spread of harmful content like hate speech and misinformation. On top, pressure from civil society and regulators to expand vague definitions of hate speech and misinformation creates a direct tension with fundamental rights such as freedom of expression and to petition authorities.

Also, the paper drew attention to the significant differences between commercial social media platforms like Facebook, Twitter, and Instagram and non-commercial social change platforms like Change.org, Avaaz, and MoveOn. Commercial platforms primarily focus on monetising content, while social change platforms view content as a means to drive political accountability and civic participation.

Regulating social change platforms should consider their size, scope, and potential harms while considering the consequences this might have on fundamental civil and political rights. Smaller platforms like Change.org may require less stringent regulation than larger platforms like Facebook. Additionally, platforms heavily relying on targeted advertising may require additional regulation to address the associated harms.

The emergence of social change platforms is closely tied to the digital era and the rise of online activism (Kaun & Uldam, 2018). Platforms like Change.org, Avaaz, and MoveOn leverage technology to connect people globally, empowering them to take collective action and amplify their voices. These platforms act as intermediaries, bridging the gap between individuals and institutions and providing a digital infrastructure for collective engagement. This move from traditional “formal accountability” to “digital accountability” mechanisms creates challenges to regulators. There is a need to explore this process further and address how the political system incorporates technology to channel societal demands from online civic participation platforms.

In the digital age, social change platforms utilise traditional social media platforms to enhance their reach and engagement. Platforms like Avaaz leverage multiple channels, including Facebook and YouTube, to mobilise users, build connections, and amplify their messages. However, moderation and maintaining consistent organisational control across platforms can be resource-intensive, and decision-makers should address this when developing new content regulations geared towards “big tech.”

Social change platforms have transformed the activism landscape, empowering individuals to participate in advocacy and shaping social change movements. Regulation of these platforms should consider their unique characteristics, content policies, and the

balance between minimising harmful content and preserving freedom of expression and the civic space (Earl & Kimport, 2011).

Further research could explore the impact of content moderation laws and regulations on social change platforms. As governments and regulatory bodies increasingly propose and enact content moderation laws, it is essential to understand how these regulations affect the operations and effectiveness of social change platforms. This research could explore how content moderation laws will impact the nature of these new “digital accountability” mechanisms and the rights of citizens to engage with policymakers. It could also investigate the challenges and implications of complying with these laws, including the potential tension between regulatory compliance and platform missions of promoting social change and civic participation.

Also, future research could examine how social change uses different strategies to maintain an open and inclusive digital civic space while adhering to legal requirements. Additionally, the research could explore the potential impact of content moderation laws on the ability of social change platforms to amplify marginalised voices and facilitate diverse perspectives, ensuring that these regulations do not inadvertently stifle or disproportionately affect social change initiatives (Klonick, 2018).

As we stated before, content moderation laws can pose legal and regulatory challenges for social change platforms, mainly non-commercial platforms. Research could investigate the legal implications and compliance burdens faced by platforms like Change.org, Avaaz, and MoveOn as they navigate content moderation regulations. This could involve analysing the legal frameworks and obligations placed on these platforms, examining the potential resource constraints and operational challenges associated with compliance, and exploring strategies for effectively addressing these challenges. Understanding the legal landscape and regulatory complexities surrounding social change platforms would provide valuable insights for platform operators, policymakers, and stakeholders in ensuring a balanced and enabling environment for social change initiatives.

By incorporating the impact of content moderation laws into the research on social change platforms, scholars can gain a comprehensive understanding of the legal and regulatory dynamics at play. This research would contribute to the ongoing discourse on balancing freedom of expression and harm mitigation, identifying potential challenges and opportunities for platforms, and informing policymakers on how to develop effective and proportionate content moderation regulations that support the goals of social change platforms while addressing concerns related to harmful content and public discourse.

References

About Us. Avaaz. Visited on July 3rd, 2023.

<https://web.archive.org/web/20230703213450/https://secure.avaaz.org/page/en/about/>

Abuse and spam. Instagram. Visited on July 3rd, 2023.

<https://web.archive.org/web/20230703222251/https://help.instagram.com/165828726894770>

Avaaz Expenses and Financial Information. Avaaz. Visited on July 3rd, 2023.

https://web.archive.org/web/20230703213925/https://secure.avaaz.org/campaign/en/avaaz_expenses_and_financial_information/?reasonstodonate

Bringing local context to our global standards | Transparency Center. Visited on July 3rd, 2023.

<https://web.archive.org/web/20230703222523/https://transparency.fb.com/policies/improving/bringing-local-context/>

Calvo, E. y Aruguete, N. (2020). Fake News, Trolls y otros Encantos. Editorial Siglo XXI, Buenos Aires.

Castells, M. (2012). Networks of Outrage and Hope: Social Movements in the Internet Age. (2nd ed.). Polity Press.

Change.org Statistics, User Count and Facts (2023).

<https://web.archive.org/web/20230703220558/https://expandedramblings.com/index.php/changeorg-facts-statistics/>

Coleman, J. S. (1988). Social capital in the creation of human capital. American Journal of Sociology, Supplement 94, S95-S120.

Community. Change.org. Visited on July 3rd, 2023.

<https://web.archive.org/web/20230703215851/https://www.change.org/policies/community>

Content regulation. Instagram. Visited on July 3rd, 2023.

https://web.archive.org/web/20230703221941/https://help.instagram.com/477434105621119?ref=igtos&helpref=faq_content

Crisis Misinformation. Twitter. Visited on July 3rd, 2023.

<https://web.archive.org/web/20230703220838/https://help.twitter.com/es/rules-and-policies/crisis-misinformation>

Earl, J., & Kimport, K. (2011). Digitally Enabled Social Change: Activism in the Internet Age. MIT Press.

Gillespie, T. (2018). Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media. Yale University Press.

Gilman, H. R. (2017). Civic tech for urban collaborative governance. PS: Political Science & Politics, 50(3), 744-750.

Green, L. (2018). Silicon States: The Power and Politics of Big Tech and What it Means for our Future. Counterpoint, Berkeley.

How does Instagram Use Artificial Intelligence to moderate content? Instagram. Visited on July 3rd, 2023.
<https://web.archive.org/web/20230703222100/https://help.instagram.com/423837189385631>

How to report violations of Twitter Rules and Terms of Service. Twitter. Visited on July 3rd, 2023.
<https://web.archive.org/web/20230703221116/https://help.twitter.com/en/rules-and-policies/twitter-report-violation>

Kaun, A., & Uldam, J. (2018). Digital activism: After the hype. *New Media & Society*, 20 (6), 2099-2106.

Kavada, A. (2012). Engagement, bonding, and identity across multiple platforms: Avaaz on Facebook, YouTube, and MySpace. *MedieKultur: Journal of media and communication research*, 28 (52), 21-p. <http://dx.doi.org/10.7146/mediekultur.v28i52.5486>

Klonick, K. (2018). The new governors: The people, rules, and processes governing online speech. *Harvard Law Review*, 131(6), 1598-1683.

Leidner, D & George, J. (2019), From Clicktivism to Hacktivism: Understanding Digital Activism. *Information and Organization*, V. 3.

Liste, L y Sørensen, K (2015). Beliefs About Democratic Values? Political Behaviour. Consumer, client or citizen? How Norwegian local governments domesticate website technology and configure their users. *Information, Communication and Society*, Vol. 18 N7.

Mackinnon, R., & Pakzad, R. (2018). Private Sector Roles and Responsibilities: Protecting Quality of Discourse, Diversity of Content and Civic Engagement on Digital Platforms and Social Media. *Governance Innovation for a Connected World*, 37.

Misuse of Reporting Features Policy | Twitter Help. Visited on July 3rd, 2023.
<https://web.archive.org/web/20230703221328/https://help.twitter.com/en/rules-and-policies/misuse-of-reporting-features>

MoveOn.org Civic Action and MoveOn.org Political Action's gift acceptance policy. MoveOn: People-Powered Progress. Visited on July 3rd, 2023.
https://web.archive.org/web/20230703214128/https://front.moveon.org/moveon-org-policy/?content=mo_nav

Nelimarkka, M.; Nonnecke, B.; Krishnan, S.; Aitumurto, T.; Catterson, D.; Crittenden, C., et al. (2014). **Comparing Three Online Civic Engagement Platforms using the Spectrum of Public Participation.** UC Berkeley: Center for Information Technology Research in the Interest of Society (CITRIS). Retrieved from <https://escholarship.org/uc/item/0bz755bj>

Pedro, M. (2015). The influence of political factors in policymakers' perceptions on the implementation of Web 2.0 technologies for citizen participation and knowledge sharing in public sector delivery. IOS Press.

Platform Terms - Meta for Developers. (2023). Visited on July 3rd, 2023.
<https://web.archive.org/web/20230703221809/https://developers.facebook.com/terms>

Press Centre: Avaaz Facts. Avaaz. Visited on July 3rd, 2023.
<https://web.archive.org/web/20230703215558/https://secure.avaaz.org/page/en/pressfaq/>

Privacy Policy and Terms of Use. Avaaz. Visited on July 3rd, 2023.
<https://web.archive.org/web/20230523193400/https://secure.avaaz.org/page/en/privacy/>

Privacy Policy. MoveOn: People-Powered Progress. Visited on July 3rd, 2023.
https://web.archive.org/web/20230703220052/https://front.moveon.org/privacy-policy/?content=mo_footer

Putnam, R. D. (2000). Bowling alone: The collapse and revival of American Community. New York: Simon & Schuster.

Rodríguez Bolívar, M. P. (2015). Smart cities: Big cities, complex governance?. Transforming city governments for successful smart cities, 1-7.

Rules and Policies. Twitter. Visited on July 3rd, 2023.
<https://web.archive.org/web/20230703220736/https://help.twitter.com/es/rules-and-policies>

Skaržauskienė, A. & Mačiulienė, M. (2020). Mapping international civic technologies platforms. In Informatics (Vol. 7, No. 4, p. 46). MDPI.

The world's platform for change. Change.org. (2023). Change.org.
<https://web.archive.org/web/20230703215033/https://www.change.org/about>

Why donate to Avaaz?. Avaaz. Visited on July 3rd, 2023.
https://web.archive.org/web/20230603015358/https://secure.avaaz.org/campaign/en/why_donate_to_avaaz/?aboutus

Williamson, A. (2011). Driving civic participation through social media. European workshop: Perspectives of Web (Vol. 2).

CIVIC COMPASS  Un programa de
CIVIC HOUSE