ABSTRACT

Tools to Manage Misinformation: Measuring the Utility of an Internet Bill of Rights and Correcting Terms and Conditions Agreements

Cody Olson Hastings, M.A.

Mentor: Leslie A. Hahner, Ph.D.

The Internet fuels more societal and technological advancement than ever before in humanity’s existence. Its birth brought simple data transfers between two entities, yet now it spans the world connecting people almost instantly. Entertaining any thought of regulation for this system to diminish current harms on the Internet is quickly met with outright rejection for fear of governmental bodies deluding the public to then usher in the destruction of free society. On the other end, lawlessness breeds anarchy when bad actors are not punished. Using the Internet today entails a bombardment of messages and manipulative means to garner attention, and we, in the United States, tolerate this. We refuse to even dream of a better system out of fear. Misinformation led to the insurrection of the U.S. capitol, almost destroying our peaceful transfer of power, perhaps the most sacred facet of our government. We must imagine a better Internet.

Tools to Manage Misinformation: Measuring the Utility of an Internet Bill of Rights and Correcting Terms and Conditions Agreements

by

Cody Olson Hastings, B.A.

A Thesis

Approved by the Department of Communication

____

David W. Schlueter, Ph.D., Chairperson

Submitted to the Graduate Faculty of
Baylor University in Partial Fulfillment of the
Requirements for the Degree
of

Master of Arts

Approved by the Thesis Committee

____

Leslie A. Hahner, Ph.D., Chairperson

____

Scott Varda, Ph.D.

____

David Bridge, Ph.D.

Accepted by the Graduate School

May 2023

____

J. Larry Lyon, Ph.D., Dean

Page bearing signatures is kept on file in the Graduate School.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

It is surreal to write this now at the conclusion of this project. I want to start off by appreciating my family. Their engagement, kindness, and patience have immensely impacted my ability to be where I am today. Rescue trips and plenty of encouragement made working consistently possible. Noble, I appreciate your patience with me and support through life to get me here to reach this accomplishment. To Erica, thank you for reading through every single word of this paper to provide feedback and the 2 a.m. opossum memes. Both were invaluable. Thank you to Dr. Grossman for sparking the interest in Communication and the conversations we have shared through the years. To Dr. Hahner and Dr. Varda, you both bring a sense of family to this wonderful program and were something I hoped I would find in coming to Baylor. Thank you both so much for fostering my interests academically and taking the time to know me personally, beyond "that guy with the furry, carpeted wall", as funny as that is. Dr. Bridge, thank you for assisting in this endeavor and guiding me as well. I want to also thank Dr. Gerber, Dr. Barrett, Dr. Damron, and Dr. Rhidenour for providing wonderful and engaging classes along the way. I recall thinking of each one of our classes together at different points in writing this thesis and appreciate all I learned in my time at Baylor. I am so blessed to have such wonderful people in my life pushing me to higher heights and want you all to know I recognize the effort you have all made to impact my life. This project is the culmination of all of these forces and I am so very grateful to each and every one of you. Thank you.

CHAPTER ONE

Exiting the Dark Ages of The Internet

Introduction

Online spaces present new challenges for communication and interaction, consequentially leading to new points of scholarly inquiry as well. Study and understanding must remain consistent and thorough to tackle the immense number of interactions and dynamics that are present in online spaces. This thesis concerns the impacts on user experience and communication that disinformation and negative content online effect. Ideally, behavior and experiences online should pose little to no harm to people wishing to participate in these spaces. We are far from this state at the given moment, but scholars and practitioners are gradually defining and refining beneficial and detrimental spaces online.

To shape the topic at hand, U.S. Internet users are afforded very few protections with the current state of the internet. This prospectus will utilize mis- dis- and malinformation (MDM) in examining an Internet Bill of Rights (IBR) as a means of protecting users, and a potential path forward to offering viable change through modifications to Terms and Conditions (T&Cs) documents. The extant frameworks I analyze demonstrate unhealthy situations for communication. Better communicative practices offer solutions to the root issues of these problematic frameworks. It is imperative to reduce ambiguity and increase understanding of the rules these frameworks operate by to promote beneficial interaction and healthy norms.

I include a thorough examination of the state and function of two Internet Bills of Rights (IBR) by analyzing the communication principles of the EU's Declaration on European Digital Rights and Principles package, along with its lasting impact. I then follow the same format with Brazil's own IBR, Marco Civil da Internet.

Reformation may offer some useful tools for improving the state of the internet in the United States. Currently, US civilians are offered rudimentary protections regarding data, such as with the search and seizure of personal information in investigations. This type of protection, while helpful, fails to adequately address the full scope of the ways U.S. civilian data is handled. Other protections largely take shape in monitoring the companies that parse through user data, rather than giving control to users. U.S. law does not grapple well with how user data is currently harvested. Our offline lives do not feature an ever-present journalist recording what we do every second of our lives. However, our phones constantly send data on our behaviors, activities, patterns, and interests. These are all tracked when using the internet. This kind of invasion of privacy is not tolerated offline and is key to the reason why an IBR is so sorely needed. We must reign in the harvesting and usage of user data. Defining and communicating our rights is only the most basic step in building a better online environment for U.S. users.

Terms and Conditions (T&C) are established primarily to avoid liability in interacting with companies. These documents are useful in clearly defining the extent of the interaction between two entities through the eyes of the law. Users are allowed to participate on Twitter only after having agreed to abide by the rules that Twitter established. Users banned from Twitter cannot, therefore, claim unfair punishment since they agreed to the rules at the start. This framing is overly simplistic to offer the best reasons why T&Cs are beneficial, navigating liability is essential. However, as time has progressed and T&Cs have lengthened, many users face undue burdens to fully read and comprehend these documents. Apple is a clear example of relevant obstacles to comprehension. Before even using their phones, users are greeted with a T&C document to read that lays out all the rights Apple reserves when you purchase their device. Again, some are beneficial to Apple maintaining profitability, but other examples are cause for alarm, such as TikTok. This application is scrutinized for the incredible amount of information that is harvested from users and sent abroad. Whether buying an iPhone or downloading applications, the vast majority of users lack the skills and education required to understand what they are agreeing to. Average citizens simply do not possess the ability to view and understand these documents in the same way that our courts do. This area is prime for communication solutions to equip civilians to maintain their privacy and digital well-being.

This thesis uses a case study approach, in that I examine each tactic for managing online communication with explicit reference to a specific instance. Case study approaches to communication are useful for observing and understanding scenarios where the "individuals, messages and context" are relevant (Sellnow et al., 2009). With the rise of MDM in online spaces, studying "how messages shape perceptions" through de-platforming, IBR, and T&C clarifies what approaches may prove beneficial or harmful (Sellnow et al., 2009). Moreover, a case study approach allows me, as the researcher, to elucidate the assorted outcomes of these corporate responses to community interactions. In what follows, I aim to highlight the pressing need for research on IBRs and T&C as responses to MDM.

The main point of pursuing these topics is to provide further insight into the issue of translation. Twitter wants its users to behave a certain way, civilians want to avoid discontinuation or ejection from online communities, and people drafting T&C used to facilitate these transactions want their terms understood. This consideration points to the additional challenge to the central issue of message translation. The content clearly needs refinement, but the issue extends further into the form of the message as well.

Literature Review

The extant literature in this arena of communication scholarship is broad. To clarify, I divide each area of the thesis and review relevant materials to demonstrate how this thesis will contribute to current research. I attempt to draw lines of connection between internet practices, platform logistics, and ultimately, the communication principles at stake.

Internet Bill of Rights

An Internet Bill of Rights may serve as an effective remedy to bridge the current gap in regulatory power between government and the social spaces offered by private companies, such as Twitter. This avenue for legislative action would mean drastic and sweeping changes to the online landscape and the companies that manage them. Because of this, this section observes the brief history of this movement and potential avenues for installing this type of system as a measure of protection for US civilians connected to the internet.

The Edward Snowden data leaks in 2014 brought attention to the government's ability to monitor and store information about civilians. Such stores of data about the population raised questions on the "legality, necessity and proportionality standards governing State surveillance powers" (Ni Loideain, 2015). Metadata emerged, as a term, to encapsulate the information that is produced "as a consequence of a communication's transmission" (Ni Loideain, 2015). What encapsulates a broad degree of information that ranges from the physical location of the communicator, the time it took place, how the message was sent, as well as more contextual information behind the actual message that was sent (Ni Loideain, 2015). This information is produced as soon as the device is switched on and leaves a "constant trail of metadata" produced by the ever-connected smart devices so commonly used today (Ni Loideain, 2015).

Metadata became the focal point for a series of legal developments within the EU. The EU Data Retention Directive, a mandatory two-year holding of civilian metadata, was struck down through "[establishing] that "data sovereignty' is a key element of the right to the protections of personal data" in the EU Charter of Fundamental Rights (Ni Loideain, 2015). Further movement for the protection of user data took shape in the new right EU citizens had to remove links from the internet where concerns of encroachment into their private lives were present (Ni Loideain, 2015). Both decisions marked stark changes to the internet, namely the permanence of data on the internet, and the ability for average citizens to have links to their private information removed from the access of others.

Further insight should be taken from Marco Civil da Internet, Brazil's own IBR. Drafted and put into law in 2014, this legislation attempted to address how regulation may grapple with an ever-changing internet, as well as "preserving fundamental rights" to avoid harming the social nature of online communication (Affonso Souza et al., 2017).

Topics on "privacy, data protection, net neutrality, liability and copyright" framed much of the legislative conversations that lead to the creation of this legislation (Affonso Souza et al., 2017). These issues have persisted and remain relevant to the conversations currently entertained over internet security in much of the US.

One point of reflection lies in how Marco Civil da Internet deals with anonymity. Many online spaces in the US feature anonymous elements, so any potential legislation for the US will need to address this feature more satisfactorily than the "blanket prohibition of encrypted channels for online browsing" that is present in Brazil's IBR. (Aftab, 2022). This can be done though, as the "European approach… [recognized] the importance of anonymity for legitimate purposes" (Aftab, 2022).

An Internet Bill of Rights for the US carries a different meaning when compared to either regulation operating in Brazil or the EU. Progress in this legislative avenue is encouraging to see, especially considering that these problems only recently have come into public focus. However, further tailoring must occur for any form of effective and meaningful protection to pass through Congress.

One avenue to consider is whether or not to apply a "heteronomous system of regulation, or self-regulation" as a system of checks and balances to the bill (De Minico, 2015). This question is better expressed when you measure how the internet changes the relationship between governing methods by leaving oversight to private entities or governmental agencies. Either choice entertains the possibility of mismanagement. Self-managing private entities, without oversight "pursue only egotistical interests" meaning "the achievement of the common good depends on… [correspondence] with private interests" (De Minico, 2015). The alternative features private entities as little more than "indirect [administrators]" of the state that lose "regulating and managing autonomy" (De Minico, 2015). Either option presents the possibility of failing to properly guarantee the rights afforded by such a bill.

This is precisely where involvement and tailoring offer a better solution than blanketing governance. De Minico determined that the regulation done to tackle the prevalence of online piracy serves as a viable legislative comparison (2015). They argue that Creative Commons licensing balanced out authors' rights to their media while also offering consumers the ability to conveniently purchase and view the newly licensed content (De Minico, 2015). The interests of both parties were protected and "a precise order between heteronomy and private law" was secured, which is exactly the type of balancing such a document must do (De Minico, 2015).

There is some consensus on what rights to include within a user's bill of rights. Most of what these proposed bills offer "appear to affirm or adapt existing rights… in the digital context" (Micheal Yilma, 2017). However, there are still points of contention. These have taken shape through two opposing viewpoints that argue whether to invent new rights or simply extend current rights into the online realm (Micheal Yilma, 2017). Italy produced its own Internet Rights declaration in 2015 that seemingly followed the latter viewpoint. Rather than invent new rights, the document expresses the "purpose of 'affirming' existing rights in the Internet context" (Micheal Yilma, 2017). These different approaches are ripe for study and distillation to aptly guide future legislation.

Despite the seemingly superficial nature of the document, it still produced notable change and protections. The legislation focused on "rights which could be called subsets of the right to privacy" to work from established precedent (Micheal Yilma, 2017). It is through these subsets that the document then legitimized a total of fourteen Articles of Declaration, ranging from the "right to information self-determination" to the "right to be forgotten by search engines" (Micheal Yilma, 2017).

Beyond legislation, there are issues with determining how best to exercise and communicate the protections afforded by these rights. The nature of the internet does not permit the same policing that we institute in offline settings. Instead, much of the obligation to see these rights respected falls onto Internet Service Providers (ISPs) as they grant and determine what is accessible on the internet (Bassini, 2019). Another facet to consider in this situation is how civilians are to voice issues should their rights be infringed upon. US citizens typically file suit with the government and go through various courts and circuits to have their issues resolved. ISPs functioning in this manner present an undue burden on these entities and an "underlying paradox… [where] ISPs should remain as much insensitive as possible to third-party content…but… feel pressure to take positive steps" to seeing online rights respected (Bassini, 2019).

Copyright is an area that exemplifies this issue well, as it requires an appropriate "balance between freedom of speech and other competing rights" (Bassini, 2019). Issues around this topic culminate at the point where ISPs remove content from the internet. Copyright and the removal of unlawfully submitted content are difficult systems to manage due to the nature of requiring "notice and takedown mechanisms" that exempt ISPs "from liability for unlawful content or activities" (Bassini, 2019). To ground the topic, fair-use and fake news both currently require review beyond what current algorithms are capable of handling and beyond what ISPs could reasonably screen (Bassini, 2019). Such a burden on private entities forbodes scenarios that place the "evaluation on the takedown of content… on ground that are most likely 'biased' towards ISPs' business interests" rather than legal reasoning (Bassini, 2019).

Greater attention and focus will allow for the crafting of solutions to these complicated issues. The lack of relevance is one of the primary hurdles to clear and is one of the aims of this thesis. At current, no concerted effort encapsulates how "international, regional [or] national human rights regimes fail to protect human rights in the digital environment" (Yilma, 2022). Current IBR initiatives show too much focus on managing challenges and fail to properly show why such initiatives are important (Yilma, 2022). IBRs are currently lofty, demonstrating "visionary but unrealistic demands" which hurt the salience and credibility of such efforts (Yilma, 2022). Solving issues first requires a mutual understanding of the issue and an agreement that it is an issue worth solving. For these reasons and the challenges stated above, this thesis will attempt to distill and surmise the issues we face in lacking IBRs as the online world grows.

Terms and Conditions

If you own a smartphone, you have likely encountered a Terms and Conditions document. These documents lay out the parameters for the interaction between the user and the entity supplying the service. On the surface, these documents are vital to providing an understanding of what the exchange will entail. The goal is to shield both parties from actions the other does not agree to entertain. However, the current state of these documents places an undue burden on average citizens that are not equipped to understand what they agree to. I intend to demonstrate this by outlining how these documents fail to provide informed consent and how various groups currently interact with these documents.

What is on the table for companies to ask of you in Terms and Conditions documents? No average person knows. The overwhelming majority of people agreeing to these documents do not read them (Luger et al., 2013). The exceptional one percent of people that do read before agreeing only spend 29 seconds scanning the document (Luger et al., 2013). This is a problem. Any illusion of users knowing what they agree to fades entirely when considering that these documents are both dense and feature complex verbiage (Luger et al., 2013). Yet the people agreeing to these documents feature reading levels often far below what is required to consent to such documents. Luger et al. (2013) found that half of the adults in the UK possess reading skills worse than 14-year-old children yet revealed that graduate-level skills were necessary to understand documents of this type.

The EU regulation defines user consent as a "voluntary, specific and informed 'indication of his wishes'" that alludes to "users… [retaining] the power to control their personal data" (Luger et al., 2013). Considering this, we must look at how average citizens engage with these documents to understand what issues are present. Luger et al. (2013) found average people read T&Cs poorly, displaying "over-emphasis upon the insignificant details within a text…difficulty in identifying the key concepts… [and] often do not consider the context of the narrative.” These findings point towards the reality that the vast majority of information in these documents is not understood, not satisfactorily consented to, and represents vulnerabilities in protecting civilian rights.

Observing how people engage with these documents yields further useful information. Though the data from Luger et al. (2013) shows that T&C engagement is not sufficient, further study revealed marked differences in groups that were presented first with a privacy policy before viewing a T&C document (Steinfeld, 2016). Specifically, participants demonstrated "more time and effort reading" compared to those that "indicate their agreement without being presented with the policy be default" (Steinfeld, 2016). Data, such as this, points to the potentiality of communication solutions providing a path forward to correcting user engagement with these documents.

Issues of informed consent only grew more relevant over time as T&Cs encroached on everyday life. Specifically, the Internet of Things (IoT) introduced a deluge of options for people to enjoy. Everything from smart coffee makers to smart doorbells served as "physical access points to Internet services" and, therefore, require agreeing to T&Cs to use. O'Conner et al. (2017) conceptualize the issue well, finding that "the ubiquitous nature of IoT… [presents] more of a challenge as data may be collected without the digital health citizen being aware". Though this pertained to the health industry, the key issue of improperly consenting to the encroaching IoT issue is not unique to this sector. The EU began the General Data Protection Regulation in 2018, which focused on the producers of these contracts following the principle of "Privacy by Design" (O'Connor et al., 2017). This regulation placed "emphasis [on] transparency, security and accountability" to better protect users from uninformed data collection (O'Connor et al., 2017).

Further technological penetration into civilian lives also entails increased interaction with more vulnerable groups. An overwhelming majority of the major social media platforms available to the EU and US do not permit children younger than thirteen to use their platforms (Schneble et al., 2021). Exemptions to this rule, though few, typically permit adults to provide consent (Schneble et al., 2021). It is therefore frustrating to see that these applications, which aim to retain their large number of users, feature differently stylized and difficult-to-read T&Cs. Schneble et al. (2019) note that though the EU's Article 29 Working Party offers "recommendations on the consent process… [they] were not able to identify a standard presentation format." This is troubling. A standardized presentation would increase user familiarity. In addition to this, a formulaic presentation allows for greater communicative refinement while still allowing private entities to include specific limitations and rules for their service. Legislation could seek a unified presentation "pictorially or in short video sequences" to properly receive informed consent (Schneble et al., 2021). This would enable a vastly wider audience to participate without undue burdens, such as the sensory impaired populations.

Current research points to the fact that the state of T&Cs do not properly engage the user to receive informed consent. This is especially troublesome when considering that vulnerable groups, such as children, use these services with few checks and balances that put the user's safety first. This is difficult to accomplish though, as children will lie about their age, and the "ease of registering for a social media service… does not constitute a barrier" (Schneble et al., 2021). Standardization should feature some form of age verification to ensure unsupervised children younger than thirteen are unable to use these spaces. Biometric data may serve as a fair solution when considering the barrier preventing the acquisition of this data is the complexity of T&Cs. (Schneble et al., 2021).

Students are also a segment of the population that interacts uniquely with T&Cs. Massive Open Online Courses (MOOCs), in particular, utilize data considered "special category or sensitive" (Khalil et al., 2018). The issue present in these organizations is that the T&Cs these students agree to are often ambiguous. These ambiguities revolve around "the range of actors… having access to the content… [are] often outside the scope and declared purpose of the initial consent" these students provide (Khalil et al., 2018). Khalil et al. (2018) then observed multiple levels student data passes through, such as student-to-teacher, student-to-institution, and lastly institution-to-institution.

Students display differing levels of awareness of the way their data is handled corresponding to how students perceive the relationship. For the sample, students felt their data is "used directly to support their own learning" (Khalil et al., 2018). Fewer students expected their data to benefit the betterment and management of the institution as a whole (Khalil et al., 2018). Why is this? It lies in the ambiguity of the T&Cs they agree to. Khalil et al. (2018) point out that some MOOCs, specifically Coursera and Iversity, inform students that they reserve the right to retain personal information if students "[use] login details from a third-party site.” The ramifications are not readily apparent with this language. Khalil et al. (2018) poignantly observes that this language includes situations where students log in from personal media sites, like Facebook, thereby granting these institutions access to personal data. This issue only expands as the scope of the relationship broadens.

At the institution-to-institution level, data is collected from students primarily to benefit MOOCs (Khalil et al., 2018). The data collected, however, is used for purposes "far beyond student learning and insights for content providers" which points to the issue of ambiguity in these institutions' T&Cs. They simply state the data's use and fail to inform students of the methods the goal is accomplished by (Khalil et al., 2018). Informed consent is attained through involvement and understanding, so using ambiguous language around data collection fails to establish consent as students are not provided a complete and clear picture of their data's use.

Methods

I plan to address these topics through four case studies guided by models of mis- dis- and malinformation (MDM). This will provide useful insight into the situations that comprise these topics and any similarities they share with MDM. These case studies will show the efficacy of current solutions to the issues listed above and where improvement is needed. This allows for further identification of potential issues for the development of effective communication-centered solutions.

Covering the EU's IBR requires more nuance to grapple with the issues it seeks to address. I use the EU’s IBR to analyze what protections it offers and how this case study may benefit a US context. To perform my analysis of this text, I will examine the EU legislation for its goals and measures to assess how well these may be met to then assess which features need further refinement to fit the US context. I then compare These documents are often lofty, so distilling the essence of what they seek to do and the measures they function by will be essential to determine how likely they are to succeed. This is one potential area where communication principles may alleviate the issue of communicating rights to civilians.

I am to answer two questions as the primary drive behind the first half of this project. Does the EU's IBR fulfill its purpose and offer users some useful pathways to manage their data? Secondly, what aspects of the EU legislation are helpful or harmful to a US context? As this area lacks critical study, this project attempts to begin the conversation into how legislation may change the rate with which misinformation spread and harms people.

Answering the first question requires critical analysis of the legislation to spot a potential weakness in language or points of ambiguity that require further refinement for actual enforcement. Grappling with the second question necessitates a contextual approach to set the stage for how rights differ in these two regions. Once that is established, I will then attempt to translate how the US may adopt a similar policy, along with the ramifications of doing so.

Assessing the issue of Terms and Conditions requires less broad engagement, so critically engaging with these documents is a narrower endeavor. As this field lacks development, I aim to address the harmful situation that currently exists from the fact that these documents fall far below the bar of providing informed consent. The burden of communication should be on the companies crafting these documents, so I will address what communication principles are useful in remedying this situation. Primarily, how can T&Cs be improved to better accomplish their goal of managing liability by informing both parties of the obligations and permissions stated in these documents?

I plan to answer this question by addressing the T&Cs of two separate communities of different standings, such as the closed ecosystem that Reddit manages with its subreddit architecture and the more misinformation-plagued environment Twitter fails to remedy. Featuring these two communities provides polar views on how communities are managed, as Subreddits utilize more crowd-sourced methods of reporting compared to Twitter. I conclude by parsing through the subreddit r/WorldNews to evaluate how well this community informs users of its policies.

CHAPTER TWO

Legislation to Combat Misinformation

Introduction

This project aims to improve the well-being of citizens online. The literature review provides an understanding of several large and growing features in US society where civilian rights and well-being are harmed and deserve greater attention for correction. Critical engagement with my topics is essential to understanding the contextual nature of each of these issues. De-platforming, IBR, and T&Cs are all broad topics worthy of deep analysis. This project will face limitations in how far my analysis can reach, so my observations and findings will reflect overall trends rather than clearly defined findings. This is still a worthy endeavor as my research will direct future scholarship in deeper engagement with these topics from the insight this project will provide.

In summary, the state of user protection in the US is remarkably low. Care and management of users are left largely to private entities that either have conflicting interests in the management of their userbase or possess inadequate means to manage the populations that utilize their services. For this reason, changes in how these entities communicate with their userbase must occur to improve online spaces so that users receive the same ethical and fair protections they enjoy in offline environments.

My literature review provided a brief introduction to two avenues where the mistreatment of US civilians is plain to see, yet actions to correct these wrongs lag from what we accept in offline settings. For these reasons, this project will contextually situate each topic to provide an understanding of the issues at hand. This will then lead to commentary and comparison of separate case studies to begin building information for future scholars to craft salient solutions in pursuit of placing the well-being of internet users at the forefront of concern and a reasonable path forward to improving these areas.

The broader movement of digital constitutionalism has seen several iterations since the cusp of the twenty-first century (Gill et al., 2015). Popular movements adapted and changed thematically over time to gain support and legislative legitimacy across the globe (Gill et al., 2015). An Internet Bill of Rights, such as the Declaration on Digital Rights and Principles passed at the end of 2022, condenses some of these broad themes in digital constitutionalism. General digital constitutionalism concerns barriers to free association on the Internet, defining the affected parties, a formalized method to seeing rights guaranteed, and often lacking applicative means of realization (Gill et al., 2015).

The EU recognizes several harms stemming from MDM content in the wake of the COVID-19 pandemic. The public's overall confidence in vaccine benefits dropped in the mid-2010s (De Figueiredo, A. et al., 2020). Additionally, several channels of influence have emerged on which the public bases their assumptions. The most to the least damaging channels of influence were from medically credentialed people stoking fears, individuals profiting off marketing vaccine arguments, the politicization of vaccines, and misinformation superspreaders (Larson, 2018). Surveys showed "Europe as the region with the highest skepticism around vaccine safety", likely leading to several specific additions to their IBR to combat some of the features of MDM spread (Larson, 2018).

European Declaration on Digital Rights and Principles

The European Declaration on Digital Rights and Principles bill stands as one of the first of its type of legislature. Previous IBR movements called for change and action, yet none possessed legislative teeth to combat MDM content, as I will demonstrate. Analyzing this document for its potential ramifications provides valuable insight for guiding future efforts in combating MDM on a national scale. The pandemic highlights MDM content's ability to bridge communities and transfer its influence beyond the internet, beyond television, and is actually able to invade even the smallest segmented communities of society. Less confidence in vaccines harms every person in society. More critically, average people cannot protect themselves against the influence of MDM, given its pervasive and evocative nature. This document is the first actualized step in combating and correcting the path malicious actors have set society-at-large down and deserves scrutiny.

Given that this legislation is quite recent, having only become law in 2022, there are no immediate metrics to understand whether this bill has changed user experience. In this light, this chapter analyzes how the language of the bill articulates goals for rights and whether the language of the bill accounts for the most pressing issues of violation of rights present in many forms of MDM. In particular, I will analyze how the language and direction of this declaration speaks to ongoing research regarding MDM and the reduction thereof.

The EU created its Declaration on Digital Rights and Principles very recently, in its attempt to govern online actions and values. Condensing these themes into actionable aims, as opposed to overgeneralized sentiment, is critical in these documents achieving measurable change in the sectors they aim to reform. The EU's declaration on these rights and principles features six chapters to address the broad themes these documents tend to concern. These are, Chapter I: Putting people at the centre of the digital transformation, Chapter II: Solidarity and inclusion, Chapter III: Freedom of choice, Chapter IV: Participation in the digital public space, Chapter V: Safety, security and empowerment, and Chapter VI: Sustainability. These chapters establish the following rights for Internet users: the right to having a person-centered online experience void of barriers preventing access, having unconstrained access to choose your method of participation in larger communities or simply as a resource of information which as both free from the risk of data theft and promotes greater participation through healthier online communities. Overall, this declaration attempts to concretize values for digital governance. As it relates to MDM, a large portion of the declaration proves exemplary for analysis.

For this case study, I primarily analyze the third, fourth, and fifth chapters, as these selections have the highest likelihood of reducing MDM. Chapter three addresses the emergence of algorithms, AI, and the interplay of these two factors in creating a fair digital environment. Chapter four concerns how users participate with online content, with other users, and the methods of content delivery. Chapter five aims to ensure online spaces are secure, that digital content is authenticated, and also maintains the protection of privacy. These selections coincide with many of the elements involved in the spread of MDM. As such, they offer the best measure of how well this document will function as a form of protection for users from the harmful content currently plaguing online information spheres.

The interplay between algorithmically promoted content and toxic interactions in online spaces is a complicated dynamic. One piece of this issue lies in how users have grown increasingly isolated from the general population online. Instead of a familiar communal exercise, online participation features increasing levels of fragmentation (Allison & Bussey, 2020). Broadly, this trend indicates that users relate less and less to each other (Allison & Bussey, 2020). Algorithms contribute to this phenomenon by suggesting similar content to each specific user. YouTube's video suggestion algorithm readily supplies users with a plethora of related content based on a single viewing of a popular video. YouTube and adjacent systems insulate users from broader topics of interest, instead sending users into increasingly niche spaces based on user engagement (Allison & Bussey, 2020). Unfamiliarity combined with the emotionally evocative nature of online content then manifests user-to-user interactions lacking care and consideration people afford each other in more communal offline experiences. This is one avenue that an IBR can seek to remedy.

The U.S. faces this issue presently, where a large segment of the population has encountered misinformation and lacked the tools to discern truth from falsehood. The propaganda surrounding the COVID-19 vaccine issue highlights this well. Parts of the U.S. population believe conspiracy theories asserting the claim that the vaccine magnetizes the blood of people who receive these injections (Fogarty, 2022). Interactions between communities that believe and disbelieve such theories feature high levels of toxicity (Pascual-Ferrá). Common tropes of these conversations involve personal degradation, among other toxic behaviors. An IBR is one possible avenue to correcting the issue with people lacking a consistent view of the factual nature of the world.

Mandating features that point users to salient, clarified information can hinder the progress of conspiracy theories thriving off of competing, unchecked viewpoints.

Community is central to the spread of MDM. Expulsion is one current solution social media adopted to punish violators of community guidelines. Literature on this tactic points to a general reduction in the spread of MDM (Jhaver et al., 2021). However, highly motivated users banned from popular platforms tend to migrate to platforms that feature lower degrees of content moderation (Jhaver et al., 2021). We do not know the lasting outcomes of banning and de-platforming users in these spaces, as this policy is relatively new. The current practice creates echo chambers that reject contradictory narratives (Ali et al., 2021). Freedom of expression is foundational to the United States. In line with this, an IBR for the U.S. should seek a similar balance to what is enjoyed in offline conversations. Namely, the digital environment must not foster systems that encourage insularity and isolation but instead promote a safe blending of differing opinions.

The topic of platforms introduces additional methods for reducing the spread of MDM. The visibility of platform hosts is one issue with user-to-content and user-to-platform interaction. Mandating transparency for who owns and operates social media sites will enable users to catalog known sources of MDM, such as Alex Jones' InfoWars platform. Granting U.S. Internet users access to know who manages the sites they visit will further bolster the general public's awareness of conspiracy theory-laden platforms. Part of the battle in stemming the flow of MDM is building a catalog of trustworthy, reputable sources for users to receive news. The identification of less reputable sources aids in this endeavor. Mandating data and platform transparency is something an IBR can accomplish.

My analysis will reveal what exactly the EU's IBR has the potential to accomplish by connecting the themes of its aims to literature's understanding of what encourages and discourages the growth of MDM. Coinciding with this, performing a critical analysis of this legislation to distill key points is critical to this paper. Metrics for evaluation concern demonstrable methods for reducing MDM in the online sphere. Namely, I will assess how user-to-content, user-to-user, and user-to-platform interactions are altered with the introduction of this legislation. Useful IBRs will feature changes that align with current literary consensus on effective methods for reducing MDM in these areas.

Chapter III: Freedom of Choice

Chapter three of the EU declaration introduces four general aims to address the use of algorithms and AI in sorting, using, and delivering data to users online. The first two are:

> - Artificial intelligence should serve as a tool for people, with the ultimate aim of increasing human well-being.
> - Everyone should be empowered to benefit from the advantages of algorithmic and artificial intelligence systems including by making their own, informed choices in the digital environment, while being protected against risks and harm to one's health, safety and fundamental rights. (*European Declaration on Digital Rights and Principles | Shaping Europe's Digital Future*, n.d.)

The first aim addresses human well-being as the central focus for what AI should enhance. This opens multiple avenues for analysis. Improving well-being is defined as a cumulative process that centers around increased awareness of AI and how these systems function online. A critical point, though, lies further down in section *b*. Promising to inform users of their interactions with artificial intelligence is an interesting addition, as artificial intelligence is a broad category. This type of language, on the whole, follows commonplace aims few disagree with. The lack of descriptive measures has kept previous IBR movements from recognizing legislative action. While the language above provides an aspirational outlook, the following inclusions give these statements more legislative legitimacy:

> a. promoting human-centric, trustworthy and ethical artificial intelligence systems throughout their development, deployment and use, in line with EU values;
> b. ensuring an adequate level of transparency about the use of algorithms and artificial intelligence, and that people are empowered to use them and are informed when interacting with them;
> c. ensuring that algorithmic systems are based on adequate datasets to avoid discrimination and enable human supervision of all outcomes affecting people's safety and fundamental rights;
> d. ensuring that technologies such as artificial intelligence are not used to pre-empt people's choices, for example regarding health, education, employment, and their private life;
> e. providing for safeguards and taking appropriate action, including by promoting trustworthy standards, to ensure that artificial intelligence and digital systems are, at all times, safe and used in full respect for fundamental rights;
> f. taking measures to ensure that research in artificial intelligence respects the highest ethical standards and relevant EU law. (*European Declaration on Digital Rights and Principles | Shaping Europe's Digital Future*, n.d.)

Writing clause *b* in this manner suggests general bots also fall under this ruling. Sections *a* and *e* combine to potentially allow massive reform to online platforms. Users online interact with bots at high rates. Even though "bots are considered less credible than humans", they still exert a "significant impact on public opinion" (Hajli et al., 2022). Equipping people with the tools they need to correctly separate human users from bots will aid in correct identification and therefore lower their degree of influence.

Current research observes social bots' abilities to "[spread] unverified information" increases with higher levels of disinformation propaganda (Hajli et al., 2022). We must look to the specific rights that are employed through the language of bringing both "trustworthy and ethical artificial intelligence systems" that are "in full respect for fundamental rights" (*European Declaration on Digital Rights and Principles | Shaping Europe's Digital Future*, n.d.) In this regard, social bots must adhere to all enumerated fundamental rights normally reserved for EU citizens yet not receive the same protections.

Article three of the Charter of Fundamental Rights of the European Union relays "the right to respect [one's] physical and mental integrity". This language sets precedent for action through the violation of the right to mental integrity in accordance with the newly established precedent requiring AI to function as trustworthy and ethical systems. Though this right traditionally protects EU citizens in the field of medicine, the right to mental integrity "stresses a person's right to control their brain states… [through] the concept of informed consent" (López-Silva & Valera, 2022). Though not specifically outlined, the spirit of informed consent appears in aim nine and garners support through clause *b*'s stance for providing identifying information to users when interacting with any AI. Aim nine coupled with clause *b* then fall in line with literature's stance to combat bots, given our understanding that bots function as a pervasive force spreading MDM content. Reducing the potential for average users to mistakenly assume bots are real humans lowers trust in the content bots post, thereby reducing the efficacy with which MDM content invades communities.

Aims 10 and 11 further elaborate how the digital environment should function:

> - Everyone should be able to effectively and freely choose which online services to use, based on objective, transparent, easily accessible and reliable information.
> - Everyone should have the possibility to compete fairly and innovate in the digital environment. This should also benefit businesses, including SMEs. (*European Declaration on Digital Rights and Principles | Shaping Europe's Digital Future*, n.d.)

The generality of these aims mirrors their corresponding committals, for they do not carry the same explicit direction as the prior committals demonstrate:

> a. ensuring a safe and secure digital environment based on fair competition, where fundamental rights are protected, users rights and consumer protection in the Digital Single Market are ensured, and responsibilities of platforms, especially large players and gatekeepers, are well defined;
> b. promoting interoperability, transparency, open technologies and standards as a way to further strengthen trust in technology as well as consumers' ability to make autonomous and informed choices. (*European Declaration on Digital Rights and Principles | Shaping Europe's Digital Future*, n.d.)

Addressing public trust within the digital landscape is essential to reducing the spread of misinformation. People suffered from low trust in public health organizations as a result of the infodemic surrounding the COVID-19 pandemic (Gradoń et al., 2021). Low trust resulted from "information overload" which leads to "misinterpretation and poor decision making" (Gradoń et al., 2021). Therefore, the aim of transparent and reliable information brought on by requiring greater disclosure on part of information providers will permit faster and easier recognition of good and bad actors.

Measures that place barriers of entry to providing information on the Internet will help reduce the volume of disinformation spread via disreputable sources. With this understanding, increasing trust in valid sources of information will also lead to better informational outcomes, given that users' "belief in the reliability of information is the strongest predictor of [sharing without verification]" (Khan & Idris, 2019). Creating a fair digital environment will lower user's trust in known disinformation actors and lower the chance of users sharing without verifying (SWV) information. Chapter three alone already provides several useful tools and directions to modifying the online landscape with measures in line with the current literary consensus on reducing MDM. This IBR accomplishes more, though, in chapters four and five.

Chapter IV: Participation in the Digital Public Place

Chapter four describes user-to-information and user-to-platform engagement. Specifically, this chapter aims to establish and define inclusive access for all participants:

> - Everyone should have access to a trustworthy, diverse and multilingual digital environment. Access to diverse content contributes to a pluralistic public debate and effective participation in democracy in a non-discriminatory manner.
> - Everyone has the right to freedom of expression and information, as well as freedom of assembly and of association in the digital environment.
> - Everyone should be able to access information on who owns or controls the media services they are using.
> - Online platforms, particularly very large online platforms, should support free democratic debate online. Given the role of their services in shaping public opinion and discourse, very large online platforms should mitigate the risks stemming from the functioning and use of their services, including in relation to misinformation and disinformation campaigns, and protect freedom of expression. (*European Declaration on Digital Rights and Principles | Shaping Europe's Digital Future*, n.d.)

These aims continue the spirit of the document, open information available to internet users. Aim 12 includes language for shaping online communication to feature participatory interaction, seen in democratically rooted dialogue, and further equips this document to manifest change within the digital landscape. Nondiscriminatory freedom to associate bolstered by accessible information on the platforms users join is further elaborated:

> a. continuing safeguarding all fundamental rights online, notably the freedom of expression and information, including media freedom and pluralism;
> b. supporting the development and best use of digital technologies to stimulate people's engagement and democratic participation;
> c. taking proportionate measures to tackle all forms of illegal content, in full respect for fundamental rights, including the right to freedom of expression and information, and without establishing any general monitoring obligations or censorship;
> d. creating a digital environment where people are protected against disinformation and information manipulation and other forms of harmful content, including harassment and gender-based violence;
> e. supporting effective access to digital content reflecting the cultural and linguistic diversity in the EU;
> f. empowering individuals to make freely given, specific choices and limiting the exploitation of vulnerabilities and biases, namely through targeted advertising. (*European Declaration on Digital Rights and Principles | Shaping Europe's Digital Future*, n.d.)

Clauses *a* through *f* introduce several critical areas of concern for establishing rights online. An appropriate balance of governance is essential to crafting a meaningful IBR. We see this concern through the inclusion of freedom of expression on the individual and media in point *a*. Point *c* addresses banning illegal content while respecting fundamental rights to demonstrate the consideration taken in crafting this document. Despite this, point *c* explicitly avoids implementing censorship through broad monitoring services. No alternative is offered for instituting these clauses, which is a point of weakness for the potential efficacy of this document.

Clauses *d*, *e*, and *f* continue the theme of reducing misinformation and online harm. Much of this harm reduction is implied, rather than explicitly defined through language lessening online manipulation. Instead, these points support inclusive measures to align the Internet more closely to the diversity seen offline within the EU. Point *f* directly mentions circumventing some issues with targeted advertising by preserving the freedom of choice without external influence. Freedom of choice is critical in promoting diversity online, as exploration occurs less when users are fed content they like.

While difficult to gauge, research into reducing MDM supports this avenue of approach. Ensuring protection for vulnerable groups and mitigating exclusionary discourse appears to benefit minority groups who are noted to “be more vulnerable to illness, …[lack] strategies to avoid illness, and experience greater burden from governmental interventions” (Myers, 2021). It is important to note that this document does not need to fix all issues with the internet, but instead feature attainable improvements to the current online climate. In this avenue, improving the inclusivity of the internet may yield greater societal benefits only hinted at in literature.

Employing measures to combat targeted advertising also serves to reduce the filtering trend users experience when dealing with advertising algorithms. Offering content to users that are not strictly based on similarities will reduce “contributions to the radicalization and division of society” we see with current algorithms (Buiten, 2022). Incorporating more cultural and communal aspects also bears semblance to discoveries made in literature. Local health agencies, for example, improve “disease surveillance [by] acting in a culturally competent manner” (Myers, 2021). In essence, approaching communities through relevant, meaningful methods promotes greater outcomes for those communities by improving how data is collected. Outcomes like this exemplify the usefulness of approaching communities through culturally competent lenses, which is beleaguered when algorithms used by advertisers create societal divisions.

Chapter IV addresses critical points to intercept MDM content. Intervening online, as shown, carries the potential to change offline outcomes as well. MDM content plays on cultural trends and tropes to excite readers. Addressing issues inherent to communities known to entertain and suffer more from MDM content through legislation like changing what advertisers can show people may have cascading cultural influences that alleviate other societal issues. Therefore, the measures and aims in this chapter are in line with current MDM reduction research. Chapter V continues this theme in promoting further change by heightening some protections in areas frequently targeted by malicious actors.

Chapter V: Safety, Security and Empowerment

Chapter five outlines specific avenues to change how user data is handled, engagement with technology, and how technology must change to better serve EU citizens:

> - Everyone should have access to digital technologies, products and services that are by design safe, secure, and privacy-protective, resulting in a high level of confidentiality, integrity, availability and authenticity of the information processed.
>
> a. taking further measures to promote the traceability of products and make sure only products which are safe and compliant with EU legislation are offered on the Digital Single Market;
> b. protecting the interests of people, businesses and public institutions against cybersecurity risks and cybercrime including data breaches and identity theft or manipulation. This includes cybersecurity requirements for connected products placed on the single market;
> c. countering and holding accountable those that seek to undermine, within the EU, security online and the integrity of the digital environment or that promote violence and hatred through digital means.

Improvements to cybersecurity do not extend solely to the products users buy and operate. Cybersecurity also extends to the digital landscape by managing which resources users may interact with on the internet, as addressed in aim 16. Disinformation websites designed to "[generate] doubt about government actions" actively harm users by feeding into conspiratorial tropes that fall within this criteria (Daimi & Peoples, 2021). Clauses *a* and *b* address this aspect by ensuring products, services, and information offered online are authentic, as is regulated with offline services. Point *c* further extends the goal of improving the integrity of the internet by reducing harm. Attempting to reduce the harm from these actors is also supported in literature as an effective measure to reduce MDM.

> - Everyone has the right to privacy and to the protection of their personal data. The latter right includes the control by individuals on how their personal data are used and with whom they are shared.
> - Everyone has the right to the confidentiality of their communications and the information on their electronic devices, and not to be subjected to unlawful online surveillance, unlawful pervasive tracking or interception measures.
> - Everyone should be able to determine their digital legacy, and decide what happens with their personal accounts and information that concerns them after their death.
>
> a. ensuring that everyone has effective control of their personal and non-persona data in line with EU data protection rules and relevant EU law;
> b. ensuring effectively the possibility for individuals to easily move their personal and nonpersonal data between different digital services in line with portability rights;
> c. effectively protecting communications from unauthorized third party access;
> d. prohibiting unlawful identification as well as unlawful retention of activity records. (*European Declaration on Digital Rights and Principles | Shaping Europe's Digital Future*, n.d.).

Malicious actors during the Covid- 19 pandemic utilized MDM in several forms, such as ransomware and cyber scams. Attacks in this form took advantage of the state of confusion from existing governmental agencies lacking sufficient resources and structures to assuage civilian concerns (Daimi & Peoples, 2021). Ransomware crippled multiple health services through stealing encrypted health information (Daimi & Peoples, 2021). Similarly, cyber scams harmed civilians in several ways, such as sewing "confusion… through identity supplantation" (Daimi & Peoples, 2021). Aims 17 and 18 explicitly state the importance of protecting user data and information from interceptive attacks, such as cyber scams. Therefore, legislation improving the state of cybersecurity and bolstering commonly attacked systems, in user-to-content interactions with ransomware and cyber scams, will reduce harm from this avenue of MDM attacks.

Beyond general cybersecurity, specific measures address one of the uniquely persistent and vulnerable populations on the internet:

> - Children and young people should be empowered to make safe and informed choices and express their creativity in the digital environment.
> - Age-appropriate materials and services should improve experiences, well-being and participation of children and young people in the digital environment.
> - Specific attention should be paid to the right of children and young people to be protected from all crimes, committed via or facilitated through digital technologies.
>
> a. providing opportunities to all children and young people to acquire the necessary skills and competences, including media literacy and critical thinking, in order to navigate and engage in the digital environment actively, safely and to make informed choices;
> b. promoting positive experiences for children and young people in an age-appropriate and safe digital environment;
> c. protecting all children and young people against harmful and illegal content, exploitation, manipulation and abuse online, and preventing the digital space from being used to commit or facilitate crimes;
> d. protecting all children and young people against illegal tracking, profiling and targeting, in particular for commercial purposes;
> e. involving children and young people in the development of digital policies that concern them. (*European Declaration on Digital Rights and Principles | Shaping Europe's Digital Future*, n.d.).

Including children in an internet bill of rights will, perhaps, yield the most powerful and lasting results to achieving a safer Internet. Structural improvements to the internet will undoubtedly provide tangible reductions in the overall spread of MDM, which is currently seen through de-platforming measures (Agarwal et al., 2022). However, structural measures fail to remedy the heart of the issue. Users are still susceptible to MDM, even if these messages are temporarily removed or otherwise obscured. Chapter Five's final aims connotate the importance of raising future generations with the tools required to navigate the digital world in a more sophisticated manner than many users currently do.

Recognizing children’s importance to the future of digital well-being incorporates real-world data observed in MDM reduction literature. Children are more likely to engage with online content than older populations and encounter more misinformative material as a result (Howard et al., 2021). Attention given through this IBR recognizes the importance of the situation as children, though not universal, often lack the capacity to “judge the veracity of the information they encounter online” (Howard et al., 2021). Children suffer from MDM as exposure to fake news damages their mental well-being by “skewing their world view” (Howard et al., 2021). Countering this trend is one crucial step to reducing the harm from pervasive MDM messages among children.

Aims 20 through 22 all include elements of a subset of digital literacy known to reduce the spread of MDM. Social media literacy is the skill to understand “how social media operate… comprehend and handle social media interactions… authenticate information published on social media… [and] critical knowledge of how content on social media is organized and produced” (Wei et al., 2023). Current literature cannot explain the full effects of social media literacy in combating MDM. However, people with higher skills in media literacy tend to correctly identify fake news more often than the average person (Wei et al., 2023). Successfully spotting fake news is an essential first step to equipping users to resist MDM online. Aim 22 connects children’s well-being to improve their media literacy and critical thinking. The EU’s IBR accomplishes this goal by promoting the importance of safety to children from content and actors who currently manipulate them online, thus allowing children to acquire media literacy skills as they grow without as many malicious influences.

Observations

Analyzing the European Declaration on Digital Rights and Principles package yielded promising insight into the inception and creation of this bill. Though this study only observed three chapters, the provisions under scrutiny possessed remarkable similarities in direction and application to actualizing change suggested by MDM literature.

The issue of MDM within social groups has only spread further and further over the course of the past several years. Misinformative actors emboldened by the pandemic exist as a detriment to the American people even today. The U.S. must take notes on the progress and efficacy of this bill in the coming decade to better understand methods of reducing the effects of these harmful messages. Freedom of speech and the right to expression are a paramount differences between the U.S. and the EU in both legislative settings as well as in cultural understandings. Future courts must weigh possible benefits to social cohesion against the detriments of limiting freedoms on the internet if the possible benefits in this IBR are actualized.

The internet was not designed to congeal into separate spaces outside the influence of the law. In this sense, U.S. civilians interact with a system that has never undergone a systemic, balanced review. Every large system the general public operates in has some type of regulatory agency promoting the health of the public as the top priority. The internet is one of the few spaces where no such agency exists. No U.S. right is without bounds. There is always a point where our civilian rights are curbed for the good of the union as a whole. U.S. civilians do not have the right to bring guns to hospitals, shout bomb in an airport, or refuse search with reasonable suspicion because we deem these activities reasonable.

As the internet grows, younger generations are brought into the fold. Permitting fears of a potential detrimental outcome to dissuade study and tests on lessening the harms perpetuated by the current state of the internet is a disservice to the American people and an unreasonable precedent to perpetuate. Regulation of parts of the internet are attainable goals. Critical resources should receive greater protection, such as health resources during a crisis, and consumers should not face constant observation and tracking from advertisers.

We need to bring the state of the internet more in line with what we enjoy offline to stem some of the division present in the U.S. from the influence of MDM. It is impermissible to avoid changes to a system allowing and encouraging civilians to believe falsehoods, such as the Covid-19 vaccine magnetizing a person’s blood. The current state of the Internet perpetuates real, observable harms that are dismissed as an acceptable trade for the enjoyment the internet provides. We should look to the harms incurred by MDM during the pandemic as an example of future crises and use this in combination with MDM-reducing literature to guide legislation in the future as the EU has done with this internet bill of rights.

Limitations

Adopting legislation from foreign governing bodies is one of the largest limitations of this case study. Understanding the different legislative nuances that exist from the centuries of separate legal development is outside the capabilities of this case study. This study observed whether the provisions within the EU's IBR have support from literature to reduce MDM or not. My findings indicate nearly every provision had literary support, either within the aim itself or supporting clauses. However, the support found in literature is not exclusive to the European context. Cultural differences and boundaries introduce the possibility that strategies for reducing MDM in one area of the world may not reduce it in the same way for the EU.

Examining pure legislation constrains this case study. Including external examples of how this legislation changes the way people use the internet would bolster some of the more speculative aims contained within this document. For example, it is difficult to assess how bots will be handled in the future when even smaller, more attainable, changes have not been observed. Options, such as reducing the amount of advertiser tracking users currently experience, will provide much more insight into MDM changes through avenues of reform like IBRs.

Time will reveal much, considering the newness of misinformation management through targeted legislation. Despite these limitations, the critiques and observations still provide value, as we now have studies to draw and compare from when observing the changes this document will bring to the EU.

Marco Civil da Internet

Brazil's Marco Civil da Internet featured different goals than what the EU aims to accomplish with their bill. This legislation distilled stakeholders' comments into three main points of interest to establish rights behind. The government sought "public engagement and participation" in crafting this IBR to represent the concerns of those affected by this bill (Martins dos Santos, 2020). General public concerns culminated in the finalized version. Provisions included “Freedom of Expression, Net Neutrality, and intermediary liability” (Martins dos Santos, 2020). Overall, these three different avenues demonstrate the public’s growing interest in their place on the internet and their desire to keep their experience on the internet free from constraining elements.

Populist legislation, such as Marco Civil de Internet, does not necessarily guarantee the targeted issues are resolved. Expert input into civilian concerns is vital to drafting good, effective legislation, as average citizens are ill-equipped to draft salient solutions to navigate the complexity inherent to social dynamics online. Numerous criticisms of the bill emerged since its 2014 ratification. Critics point to permitting total freedom of speech, as this bill aims to achieve, in actuality “contributes…to new forms of oppression, such as virtual bullying… [and] the propagation of hate” (Schreiber, 2022). In addition to freedom of speech, this IBR defines the way the Brazilian government functions with issues pertaining to “data privacy and liability of Internet service providers” (Marco Civil English Version, n.d.)

Marco Civil da Internet contains five chapters with a total of 32 articles spread throughout. The official translated edition of the bill describes these chapters as, Chapter I – Preliminary Provisions, Chapter II – Rights and Guarantees of Users, Chapter III Provision of Connection and Internet Applications, Chapter IV – The role of public power, and Chapter V – Final Provisions. I will examine specific articles from each chapter, as there are inclusions throughout each chapter which relate to the spread of MDM. Based on current research, I aim to reveal the interplay between the Brazilian government achieving these goals and how the provisions encourage or reduce the spread of MDM. Coupled with this, I will demonstrate some differences between this document and the necessary changes present in the EU's IBR to protect civilians against newly emergent threats online.

Chapter I: Preliminary Provisions

The first chapter in this IBR contains six total articles to introduce the formulation and function of the idealized internet, how users should engage with this system, and definitions for separate points users interact with in accessing the internet. Articles two, three, and four are germane to include in assessing the handling of MDM within this IBR:

- The discipline of internet use in Brazil is founded on the basis of respect for freedom of expression, as well as:
  I – the recognition of the global scale of the network;
  II – human rights, personality development and the exercise of citizenship in digital medias;
  III – plurality and diversity;
  IV – openness and cooperation;
  V – free enterprising, free competition and consumer protection;
  VI – social purpose of the network.
- The discipline of internet use in Brazil has the following principles:
  I - guarantee of freedom of speech, communication and expression of thought, in accordance to the Federal Constitution;
  II – protection of privacy;
  III – protection of personal data, pursuant to law;
  IV – preservation and guarantee of network neutrality;
  V – preservation of stability, security and functionality of the network, via technical measures consistent with international standards and by encouraging the use of best practices;
  VI – the liability of the agents according their activities, pursuant to the law;
  VII – preservation of the participative nature of the network;
  VIII – freedom of business models promoted on the internet, provided they do not conflict with the other principles set out in this Law.
- The discipline of internet use in Brazil aims to promote:
  I –the right of all to access the internet;
  II – the access to information, to knowledge and participation in the cultural life and in the handling of public affairs;
  III – the innovation and the stimulus to the broad diffusion of new technologies and models of use and access;

> IV – the adoption of open technology standards that allows communication, accessibility and interoperability between applications and databases. (*Marco Civil English Version*, n.d.)

Introducing these concepts as central functions of the internet depicts the future Internet as a liberalized environment, where users freely access and navigate through the online world without constraint from structural barriers. From an MDM perspective, provisions for internet use combined with features such as openness, free enterprise, making the internet participative, and freely flowing communication manifest ideal conditions for the spread of MDM. We have a healthy understanding of how users operate within unregulated social spaces online. The free market of the internet in the U.S. serves as a useful example. Liberal places, which only aim to create a level playing field, fail to protect users from the divisive aftereffects of sensationalist media (Kolson, 2023).

Free enterprise entails some information services and social platforms failing, while others rise according to their ability to retain a marketable audience. Given that people “are more likely to share ideologically compatible messages”, provisions only guaranteeing basic principles instead of detailed provisions to equip users to navigate online interactions will likely fail to accomplish sections II and III of article two (Stein et al., 2023). IBRs must accomplish more than bringing offline rights to online spaces. Preserving diversity and maintaining cooperation between users online requires a greater degree of protection afforded to users online, as information spreads more rapidly in this space. Offline interactions are not immune to the spread of misinformation, but the effects of MDM content are less severe offline, since information travels less pervasively.

The rapidity with which online spaces exchange information is one key difference warranting greater protection. IBRs must, instead, adapt our offline freedoms to the different dynamic online spaces feature as “segregated networks [exhibit] a greater prevalence of misinformation” (Stein et al., 2023). Though offline spaces feature niche and segregated communities, information and influence spread online much more rapidly and have farther reach, given the unique nature of instant and global communication. Observing such linguistic differences and subtle changes to newer IBRs indicates the inability of this type of document to satisfy its stated goals. This is further supported by research promoting safeguards as communal regulation (Lazer et al., 2018).

Chapter II: Rights and Guarantees of the Users

The guarantees afforded to users are represented through two articles, and demonstrate the fixation to uphold freedom of speech and privacy rights:

- The access to the internet is essential to the exercise of citizenship, and the following rights are guaranteed to the users:
  I – inviolability of intimacy and private life, safeguarded the right for protection and compensation for material or moral damages resulting from their breach;
  II – inviolability and secrecy of the flow of users’s communications through the Internet, except by court order, as provided by law;
  III – inviolability and secrecy of user’s stored private communications, except upon a court order;
  IV - non-suspension of the Internet connection, except if due to a debt resulting directly from its use;
  V – maintenance of the quality of Internet connection contracted before the provider;
  VI – clear and full information entailed in the agreements of services, setting forth the details concerning the protection to connection records and records of access to internet applications, as well as on traffic management practices that may affect the quality of the service provided;
  VII – non-disclosure to third parties of users’ personal data, including connection records and records of access to internet applications, unless with express, free and informed consent or in accordance with the cases provided by law;
  VIII – clear and complete information on the collection, use, storage, processing and protection of users’ personal data, which may only be used for the purposes of: a) justifys its collection; b) are not prohibited

by law; and c) are specified in the agreements of services or in the terms of use of the internet application.

IX – the expressed consent for the collection, use, storage and processing of personal data, which shall be specified in a separate contractual clause;

X – the definitive elimination of the personal data provided to a certain internet application, at the request of the users, at the end of the relationship between the parties, except in the cases of mandatory log retention, as set forth in this Law;

XI – the publicity and clarity of any terms of use of the internet connection providers and internet applications providers;

XII – accessibility, considering the physical, motor, perceptive, sensorial, intellectual and mental habilities of the user, as prescribed by law;

XIII – application of consumer protection rules in the consumer interactions that take place in the internet.

- The guarantee to the right to privacy and freedom of speech in the communications is a condition for the full exercise of the right to access to the internet.

  I – cause an offense to the inviolability and secrecy of private communications over the internet; or II - in adhesion contracts, do not provide an alternative to the contracting party to adopt the Brazilian forum for resolution of disputes arising from services rendered in Brazil (*Marco Civil English Version*, n.d.).

These articles firmly situate the value of private communication yet affords little else to internet users. With that said, the spread of MDM messages and the degree of privacy users enjoy online introduces a much larger conversation. For malicious actors, privacy obscures intent and behavior. Social bots, for example, are a useful tool to spread information online aided by anonymity. They inform and interact with users in many ways but can also masquerade as real people. Very simply, they “put forth or respond to content in specific ways” in deployer-defined settings (Pagoto et al., 2019). The Covid-19 pandemic highlighted their ability to “flood [conversations] on particular health topics” often to the detriment of the conversation and audience (Pagoto et al., 2019). Cases of social bots’ influence do not translate to offline settings well and are indicative of the need to formulate new rights and protections for users online.

Strengthening the veil of anonymity between users online likely increases MDM, as users possess fewer tools to identify bots. In addition to this, the broad protective language within Marco Civil da Internet, such as obscuring “the flow of users’ communications through the Internet” hinders the ability to understand social bot behavior by monitoring how these programs function in online spaces (Marco Civil English Version, n.d.)Language lacking boundary definition, though aimed to benefit internet users, harms the digital landscape by establishing rights and protections for entities that do not deserve rights afforded to humans. Currently, AI removes bots by monitoring how these accounts operate and detecting behavior typical of bots. Disallowing AI to aggregate user data allows bots to spread more rapidly and strips users from protection against entities known to “lure people to websites with false information” fails clause IV of Art. 2o in Chapter I of this bill (Celliers & Hattingh, 2020).

Including a section for the rights and guarantees of users yet limiting the enumerated rights to two clauses demonstrates notable shortcomings in this IBR. It denotes different political and social climates, given the open-forum style of debate for how this document manifested. Thousands of user comments should have generated more substantial provisions than what this document represents. The discourse surrounding the political movements of 2014 demonstrated “the absence of defined pleas… and diffuse feeling of dissatisfaction and rebellion than properly around practical objectives” (Schreiber, 2022). There are no provisions encouraging harmonious dialogue, inclusions of the viewpoints or voices of minority audiences, preemptive protection from automated malicious systems, or recognition of any other threat users face online.

Privacy, and recourse for violation of that privacy, stand as the primary objective of Marco Civil da Internet, yet the bill falls short in detailing and executing this objective. Shortcomings indicated within this bill become especially evident when compared to newer IBRs, such as the EU's new IBR. Chapter II demonstrates the shortcomings in drafting legislation primarily through public discourse. MDM literature, largely, does not support the approaches taken in this section. Expert commentary provides saliency for the aims of the general populous. Brazil's people desired freedom on the internet but failed to conceptualize pure freedom leading to people wielding a "*mere semblance of freedom*… without needing to respect rules established in the interest of society as a whole" (Schreiber, 2022).

Chapter III: Provision of Connection and Internet Applications

Chapter three addresses one new right, empowering users to have their private information removed from application hosts, such as Twitter and Facebook. This chapter also clearly defines which entities are liable for illegally posted content and shields internet service providers from the content generated on the internet:

- The provider of connection to internet shall not be liable for civil damages resulting from content generated by third parties.
- In order to ensure freedom of expression and prevent censorship, the provider of internet applications can only be subject to civil liability for damages resulting from content generated by third parties if, after an specific court order, it does not take any steps to, within the framework of their service and within the time stated in the order, make unavailable the content that was identified as being unlawful, unless otherwise provided by law.
- The internet application provider that makes third party generated content available shall be held liable for the breach of privacy arising from the disclosure of images, videos and other materials containing nudity or sexual activities of a private nature, without the authorization of the participants, when, after receipt of notice by the participant or his/hers legal representative, refrains from removing, in a diligent manner, within

> its own technical limitations, such content (*Marco Civil English Version*, n.d.).

Compelling social media companies to remove user-generated content enables users in the U.S. and EU to reduce their presence online as well as remove personally harmful content. Though relatively obscure in 2014, deepfake technology "[gave] rise to apps… that allow users to create their own deepfakes" of themselves as well as other individuals (Kirchengast, 2020). This technology allows people to create convincing images and videos of others saying or doing things that they have not done. Deepfakes demonstrably "incite political deception, voter manipulation, commercial fraud" as these deceptive videos are shared through social media (Kirchengast, 2020). Part of the reason behind deepfakes efficacy in the harm they cause stems from a user's inability to discern deepfake from genuine videos. Generally, spotting a deepfake is "essentially a coin toss" irrespective of the amount of "social media usage and knowledge of deepfakes" (Lovato et al., 2020).

Despite seemingly offering a promising tool to users online, a key distinction in Art. 21 prevents this from functioning effectively. Specifically, this article incurs penalties on application providers only if they fail to remove content after receiving a notice. Article 21 remains ambiguous in what range a "diligent manner" ascribes to social media companies (Marco Civil English Version, n.d.)

In addition to the ambiguous language, requiring notice to act incurs natural bureaucratic delays. Requiring notice for removal effectively nullifies the usefulness of this right. Notice entails deepfakes acquiring enough attention to warrant removal. This defeats the purpose of removal as content like deepfakes incurs harm through notoriety.

One of the US's leading methods for combating deepfakes involves training AI to detect deepfakes before they amass widespread attention. These programs boast impressive results, with the technology needing "less than 2 seconds of video" to predict deepfakes "with an accuracy greater than 97%" (Güera & Delp, 2018). In addition to demonstrating an incredible ability to correctly identify deepfakes, AI more proficiently scans online content for potential deepfakes before they acquire notoriety as a benefit of existing as a program rather than a human. The human factor is a critical component of issues with MDM online. On the other hand, AI has the potential to both damage and correct discourse.

AI are not only utilized by companies hosting these platforms, as malicious actors continue to utilize automated systems, so it behooves legislators to utilize similar tools in combating these threats. The efficiency and persistence of AI are crucial features to their efficacy of spreading misinformation, so combatting these tools is best suited to other AI. Marco Civil da Internet's approach to removing harmful content via petitioning social media is not supported as an effective route to reducing MDM content and likely emboldens MDM actors, as the burden of reporting falls on average users.

Chapter IV: The Role of Public Authorities

Chapter IV addresses the future functioning of the state and its duties in interacting with the digital landscape. Though primarily serving functionally, some inclusions within this section address elements of MDM:

- The following are guidelines for the performance of Federal Government, States, Federal District and municipalities in the development of Internet in Brazil:

I – establishment of mechanisms of governance that are multi-stakeholder, transparent, cooperative and democratic, with the participation of the government, the business sector, the civil society and the academia;

II – promotion of the rationalization of management, expansion and use of the internet, with the participation of Brazilian Internet Steering Committee (CGI.Br).

III - promotion of rationalization and technological interoperability of eGovernment services, within different branches and levels of the federation, to allow the exchange of information and speed of procedures;

IV – promotion of interoperability between different systems and terminals, including among the different federal levels and different sectors of society;

V – preferred adoption of open and free technologies, standards and formats;

VI – advertising and dissemination of public data and information in an open and structured manner;

VII – optimization of network infrastructures and promoting the implementation of storage, managing and dissemination of data centers in the country, promoting the technical quality, innovation and the dissemination of internet applications, without impairment to the openness, neutrality and participatory nature;

VIII – development of initiatives and training programs for internet use;

IX – the promotion of culture and citizenship;

X – provide public services for attending citizens in an integrated, efficient and simple manner and through multichannel access, including remote access.

- The internet applications provided by public governmental entities ought to aim at:

I – compatibility of e-government services with multiple terminals, operating systems and applications for their access;

II – accessibility to all interested users, irrespective of their physical and motor skills, perceptual, sensorial, intellectual, mental, social and cultural characteristics, respected confidentiality and legal and administrative constraints;

III – compatibility with both human reading and automatic processing of information;

IV – easy understanding of egovernment services,

V – strengthening social participation in public policy.

- The compliance with the constitutional duty of the State in providing education at all educational levels, includes integrated training and other educational practices, for safe, conscious and responsible use of the internet, as a tool for the exercise of citizenship, for the promotion of culture and for the technological development.
- Public initiatives to promote digital culture and promote the internet as a social tool shall:
  I – promote digital inclusion;
  II – seek to reduce gaps, especially between different regions of the country, regarding the access and use of information technology and communication;
  III – promote the production and dissemination of national content
- The State must periodically seek to develop and promote studies, as well as set goals, strategies, plans and schedules for the use and development of the Internet in the country (Marco Civil English Version, n.d.).

Chapter IV brings promising additions for improving the state of the internet. Most of the new aims offer precedents for changing the way the government involves itself with the internet and establishing a reliable and reachable service for Brazilian citizens. Article 24’s eighth clause, though short, demonstrates some level of nuance for joining the internet. This clause recognizes the need for training programs to equip all users with some base level of understanding to navigate and operate online. However, it does explicitly state the initiatives of the state’s design for teaching people how to use the internet. Taken broadly, educating the populous and empowering them with the skills and knowledge required for internet use falls under the definition of digital literacy (Reddy et al., 2020). As demonstrated in the previous study, such aims are worthwhile within MDM management as a means of increasing the abilities of individuals to discern truthful information in the content they see online (Ali & Qazi, 2022).

In addition to digital literacy, this chapter includes directives to improve the diversity of the internet space by eliminating any and all barriers to those interested in engaging with the internet. Literary work already demonstrates homogeneity in thought reduces a group's inclination to challenge incoming information (Nikolov et al., 2021). Though not directly supported through this legislation, efforts to improve the diversity within the internet then contribute to reducing the susceptibility of communities in this space. However, as noted previously, this bill contains no provisions against hateful or harmful behavior. Because of this, the internet may still deride minority voices, unfortunately leading back to general homogeneity and susceptibility to misinformation. As such, MDM literature loosely supports the direction of this chapter in terms of education, but research does not support the total elimination of barriers to participating on the internet as a method of reducing the influence of MDM content.

Chapter V: Final Provisions

Most of the articles in this last chapter entail the legalities emergent from this legislation's instatement, but one article stands out quite differently from articles 30 through 32:

- The user shall have free choice in the use of software in his/hers own device to enforce parental control over content that the user understands to be improper to his-hers minor children, to the extent that the principles set forth in this Law and in Law No. 8,069 of July 13, 1990 are respected (Marco Civil English Version, n.d.).

The last article in Marco Civil da Internet stands out as an oddity for its place within this section. Chapter V mandates users have some form of content control setting within their devices and details how these new policies fit in with prior precedents. Article 29 mandates any device users interact with must possess parental controls to limit children's exposure to potentially harmful content online. Literature observing how children interact with MDM material is sparse, but some preliminary findings indicate children "act upon being exposed to fake information even when they do not trust the source" (Dumitru, 2020). Equipping parents with content controls benefits children as they then experience less misinformative material on the internet while they grow. MDM literature recognizes the impressionable nature of children and the importance of having tools to protect their children from harmful content and is therefore inline with reducing the influence and harm from MDM content.

Observations

Marco Civil da Internet heavily emphasizes maintaining an internet free of influential features that may direct people to form opinions from external sources, rather than from potentially malicious actors. While promising, the included articles fail to address large issues with this goal. Much of the groundwork for accomplishing a free internet manifests through placing incredible emphasis on privacy rights, the rights of user data, and defining state influence. A free internet can manifest in several ways. There are spaces that seek to permit any and all human behavior and spaces that limit behaviors known to reduce the freedoms of others. This IBR seems to comply with the former rather than the latter, given the lack of protectionary measures within this document.

Observing these articles yielded some potential routes to reducing MDM. The temporal aspect of these case studies presents unique findings, such as seeing pieces of this IBR in future legislation. The EU clearly saw problematic issues with how Marco Civil da Internet changed the online landscape but chose to improve on some of the elements of this legislation, such as increasing the emphasis on digital literacy. Establishing rights on the internet through IBRs is a unique challenge. Inventing new rights carries potentially dangerous precedent, but failing to do so and under-equipping people to navigate the internet is not without pitfalls.

Limitations

Culture is an immediate limiting factor in this case study. Different spoken languages, meanings, and values all combine to make evaluating this bill difficult to analyze from an MDM standpoint. Despite this, differences do exist between this bill and the EU’s new IBR and the examination of those differences are the main points of this study, so my findings still present value to future research over these areas. Persuasive malicious messages to Brazilian audiences may not find the same influential power over North American audiences with different cultural motives and histories. MDM may take shape differently across different cultures. Translation, though official, also entails some form of verification which may not account for how native speakers read this document when compared to English speakers reading the English translation. With this understanding, some of my critiques of this bill may lack saliency in how these policies function within Brazil’s government and audience.

CHAPTER THREE

Potentiality of Terms and Conditions to Improve User Experience

Introduction

Terms and Conditions (T&Cs) documents are almost inevitable to anyone joining the internet. These documents govern interactions far and beyond what may simply come to mind for most individuals, such as buying an iPhone, moving your digital content, and making the phone yours. Elements present within these documents can include limitations for user interaction, defining company liability regarding the use of the device they offer, and the extent of data harvesting from users. As we gradually adopt more technology, it integrates further into daily life, defining boundaries and use cases, as these documents do, will likely become a commonplace experience for most people. Smart devices were not available twenty years ago, yet now the T&C agreements featured in these purchases have extended into interfacing with a plethora of everyday products, such as new coffee makers or room fragrance devices.

Interfacing with another entity where either party wishes to diminish or eliminate liability now likely entails some form of a written agreement establishing boundaries of acceptable behavior in the interaction. For Apple, their agreement establishes rules buyers automatically agree to upon using the phone. Their Terms and Conditions agreement outlines general use, property rights from the software on the device, and a multitude of other technicalities to ensure a smooth experience for Apple and its audience. Generally, the inclusions within T&Cs are written by lawyers for interpretation through the law, so it is not uncommon to see these documents presented to ordinary people in poorly communicative forms. Apple introduces the entirety of its T&C through the home-screen menu of its handheld devices, where users are freely able to utilize swiping speed to browse to the bottom of the agreement to accept the terms and quickly bypass the screen. It is odd to see the public engage with these documents in this manner, as so few in the population possess the education to read and understand these documents.

Unsurprisingly, most people do not read or understand T&Cs (Steinfeld, 2016). Beyond this, average citizens within the U.S. do not possess the education to understand the words these documents use or the general meaning constructed within (Steinfeld, 2016). In studying how people engage with these documents, Steinfeld found nearly eighty percent of readers agree to privacy terms without opening the policy to read it (2016). People appear to regard such documents as inconsequential to the process of engaging with the many services on the internet, such as YouTube or Twitter.

Implementing communicative practices introduces the opportunity to improve the way people interact and understand such documents. Even simple changes, such as automatically presenting users with the policy instead of requiring users to follow a link to the policy significantly. Requiring users to click a link to find the policy demonstrated “significantly less effort in reading the document,” when compared to systems that automatically presented users with the document (Steinfeld, 2016). In this case, even simple changes alter the way users engage with these important documents.

Changing the user experience is critical to this case study. People must have the ability to understand what they agree to when signing these documents. Over time, T&Cs

have become more complex and lengthier depending on the application. Greater integration into average life entails further restrictions and requirements on part of the service provider to avoid unacceptable interference in peoples' lives. Not all apps place consumer interest at the forefront, however. TikTok is now infamous for the level of data it acquires from its users. In their T&Cs, using the app permits TikTok to transfer data regarding your phone's software, hardware, usage analytics, battery state, filenames, and more to their data servers for the purpose of improving their app experience. As it stands now, users do not understand the level of intrusion they grant solely by using this application. Despite this, average users can gain an understanding of these documents with communication-centered tools.

The central point of this case study is to demonstrate how current, popular forms and presentations of T&Cs lack features to translate messages within these documents to average people, that in turn manifests conditions within social media that augment the spread of MDM. In essence, people navigating online interactions guided by rules contribute less to spreading MDM when compared to people who do not. By extension, this means a more digitally literate and informed public that understands the content they find online more proficiently.

Social media platforms cannot bear the burden of educating every member of their audience, but Twitter can establish some form of transmissible messaging to inform its audience of Twitter's rules. The public, when informed and aware of the rules they agree to follow, will better adhere to the standards and rules of online platforms. People are capable of following rules in integrated systems and have done so in many ways in their lives when these rules are seen as legitimate and moral (Tyler, T., & De Cremer, D., 2009). Online platforms, though more impersonal, share many features with offline systems. Pervasive discord rampant within online social media platforms does not always follow into offline systems, for people utilize general understandings of rules and norms required for effectively participating in these offline systems.

Transit is one of the largest offline systems, with millions of people participating each day. This system functions remarkably well, despite the vast number of people on the road. People follow rules and standards required for participation. Accidents do happen, but the wider system functions as well as it does by equipping people with the knowledge and tools required to participate. Air travel, gas stations, and grocery markets all function similarly. The general public learns the rules of participation and abides by them for the benefits these systems offer when everyone buys into it. Violators of these systems are removed and, crucially, the wider public agrees with this policy. Protests do not occur for speeding tickets because the public understands the necessity of adherence to the rules, which follows for the other systems as well through a refusal of service.

Social media operates more cohesively when users understand and abide by the rules they agreed to follow when joining online communities. Online platforms present multiple avenues for comparison, as moderated spaces exist in multiple forms. Broader spaces, such as Twitter, feature site-wide rules in their T&Cs. Smaller spaces function similarly but with different presentations. Reddit’s structure allows users of the site to moderate their own subcommunities in the form of subreddits. These spaces vary widely in moderation styles but tend to feature moderation in three distinct forms: auto-moderation via bots, user reports, and communally-appointed human moderators (Iqbal et al., 2022). Though smaller in scale, these subreddits utilize “content moderation mechanisms… [to] flag and potentially remove… content" which violates community guidelines (Iqbal et al., 2022). Subreddits effectively present miniature T&Cs to new users joining these subreddits. Operating under this paradigm then permits worthwhile observation for determining how audiences change when presented with a transmissible set of rules and expectations, for online communities operate more cohesively when users understand and abide by the rules they agreed to follow when joining online communities.

Current literature finds subreddits with rules feature "significant content moderation activity" and display notable differences when compared to subreddits without rules (Iqbal et al., 2022). Further analysis revealed that human-moderated posts accounted for the vast majority of moderated action within these subreddits (Iqbal et al., 2022). Moderating actions do not solely prove rules help reduce MDM by improving subreddit cohesiveness. Instead, the types of rules users read matter. Toxic content, including fake news defined by Iqbal et al., "can be reduced by majorly designing… more robust policies for human [moderation]" (2022). Additional rules and tools may generally reduce MDM in these spaces and also in times where "the amount of data overshadows the availability of [rule]" crowdsourced moderation offers (Iqbal et al., 2022).

I will evaluate the efficacy of informing users of the rules they must follow in clear and concise means through the following case studies. This will be performed by observing and comparing differences in accessibility and readability of the rules outlined in T&Cs users agree to before joining. Social media lacking rules should display higher rates of MDM spread than spaces with clearly presented and transmissible rules. Doing so provides valuable insight into effective communication-centered means for building and maintaining healthier and less harmful communities online. Twitter will contrast Reddit's r/WorldNews subreddit as an example of an online community with high amounts of continual spread of MDM content, exemplified by the 2016 election conspiracy and the amound of medical misinformation regarding the COVID-19 pandemic present on Twitter (Bovet, Alexandre, and Hernán A. Makse 2019; Sharma, Karishma, et al. 2020)

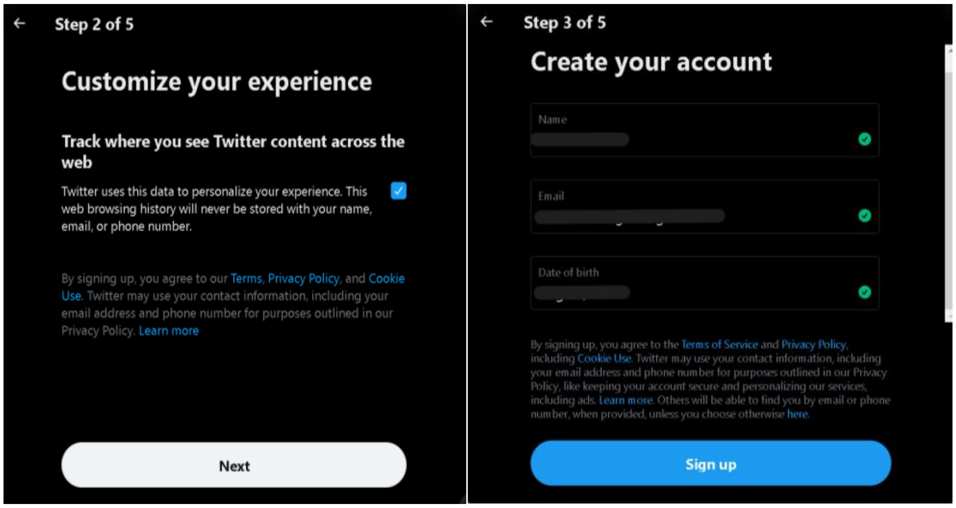
Scholarly work into misinformation continues to expand in breadth and depth over time and has so far mapped how MDM functions online through the spread and adoption of information. Despite this, Terms and Conditions documents, specifically, have not received academic interest regarding their ability to influence how MDM spreads online. Literature on T&Cs generally observes the complex nature of these documents and the various forms in which people engage with these documents (Luger et al., 2013). Scholarly inquiry into the T&Cs of social media yields more promising observations into how children sign these documents, despite the law offering a higher standard of protection compared to adults (Critchlow et al., 2020; Creswick et al., 2019). Little to no research exists connecting the readability of T&Cs to how signees may use or not use these conditions. Filling in this gap in literature potentially offers an additional useful tool for reducing MDM through more effectively communicating rules and norms to people joining online communities.

Public and private institutions alike inform the general public of acceptable behavior to perpetuate smooth interaction. Grocery stores direct customers through queues, driving a vehicle requires earning a license, and entering hospitals require multiple levels of authentication and direction to ensure continual and safe operation. U.S. businesses must inform their customers of the dangers of doing business when certain thresholds are met, based on the assumption of risk. Grocery stores are typically not hazardous places, yet skydiving is exceedingly dangerous and requires a much higher standard of information provided to customers. Information on the harms of using and participating in social media continues to expand, and the harms perpetuated in and by social media become more clearly observable and proven over time. Trends like these slowly shift social media perception from harmless places to spaces where potentially detrimental interactions occur frequently enough to deserve warning users.

Twitter

Social media platforms do not adequately inform their users of these risks to health and safety. Hazardous businesses navigate liability through mandatory informational sessions. In these sessions, such as earning a license to SCUBA dive, people are educated on the risks and informed of the specific hazards unique to diving, including how to communicate underwater, how to ensure your own safety, and how to ensure the safety of others. The ease of joining social media platforms belies the risks people undertake in joining online communities.

Twitter requires only five steps to create an account. You first provide your name, email, and date of birth, permit or decline Twitter’s request to track your web history, confirm your information, verify your email, and create a password for your account. After which, you select a picture for your profile, indicate and refine your interests, and end the process by following at least one suggested account to join Twitter. The streamlined process demonstrates the effort expended to make joining Twitter easy for average people. It is simple and straightforward to create an account and begin viewing content similar to the user's interests. However, the process in its entirety does not educate new members, in any way, about the policies Twitter holds regarding allowable behavior. Instead, these pieces of information are subtly included at the bottom portion of only two of the five sign-up panels as shown in figure 3.1.



Step 2 of 5

Customize your experience

Track where you see Twitter content across the web

Twitter uses this data to personalize your experience. This web browsing history will never be stored with your name, email, or phone number. ☑

By signing up, you agree to our Terms, Privacy Policy, and Cookie Use. Twitter may use your contact information, including your email address and phone number for purposes outlined in our Privacy Policy. Learn more

Step 3 of 5

Create your account

Name: [redacted]

Email: [redacted]

Date of birth: [redacted]

By signing up, you agree to the Terms of Service and Privacy Policy, including Cookie Use. Twitter may use your contact information, including your email address and phone number for purposes outlined in our Privacy Policy, like keeping your account secure and personalizing our services, including ads. Learn more. Others will be able to find you by email or phone number, when provided, unless you choose otherwise here.

Fig. 3.1 Screenshot of Account Creation on Twitter with T&Cs. Personal photo.

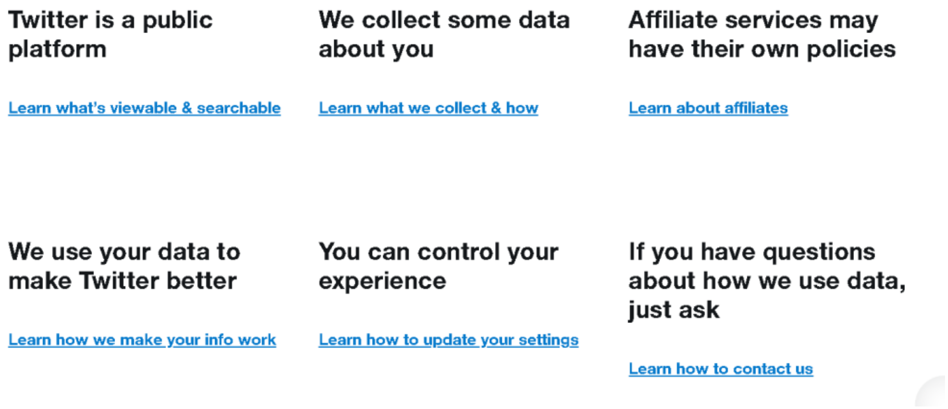
The layout of these information panels directs user attention and discourages proper understanding of terms and conditions. The bold white text contrasts with the black background and introduces the purpose of the page by directing users to fill in the required information. Distinct stylistic choices separate the text into different levels of significance. Twitter appears to want users to see the purpose of the page and how to progress through the process first, yet a key distinction emerges as users commonly scan the page from the top to the bottom. The text in the first image gradually decreases in size and darkens in color. Twitter automatically includes the option to let Twitter track where you see "Twitter content across the web", as seen in the first image. New members must manually deselect this option to avoid tracking, yet Twitter attempts to direct attention past this option. Twitter encourages users to the next, most notable, element on the page, the next button. Stylizing the page in this manner pushes users past viewing the T&Cs, privacy policy, and use of cookies by introducing these hyperlinks in small, harder-to-read text, which requires a higher degree of interest to investigate. If Twitter wanted new members to read the rules of the website, then Twitter should present them similarly to the other directive text.

Twitter demonstrates little interest in encouraging users to read their T&Cs. Requiring users to click an additional link to a separate webpage functions as an unnecessary barrier to learning how to behave within Twitter. Following these links, three separate informative pages open and describe each policy in detail. The page on Twitter's privacy policy demonstrates a more concerted effort into making the page accessible to average readers. Choices like this demonstrate some level of awareness of the concerns of Twitter users. Twitter shows it has the capacity, through its privacy policy, to make its pages and information digestible, when needed. The contrast between the readability of the privacy policy and the more mundane T&C and cookie policy hint at the heightened public concern and pressure for Twitter to more effectively communicate the nature of how user data is handled.

Twitter includes a short, stylized notice to draw readers to read before continuing further into their policy as depicted in figure 3.2.

> Before you scroll, read this
>
> It's really hard to make everyone happy with a Privacy Policy. Most people who use Twitter want something short and easy to understand. While we wish we could fit everything you need to know into a Tweet, our regulators ask us to meet our legal obligations by describing them all in a lot of detail.

> With that in mind, we’ve written our Privacy Policy as simply as possible to empower you to make informed decisions when you use Twitter by making sure you understand and have control over the information we collect, how it’s used, and when it’s shared.
> So if you skip reading every word of the Privacy Policy, at least know this:



Twitter is a public platform

Learn what's viewable & searchable

We collect some data about you

Learn what we collect & how

Affiliate services may have their own policies

Learn about affiliates

We use your data to make Twitter better

Learn how we make your info work

You can control your experience

Learn how to update your settings

If you have questions about how we use data, just ask

Learn how to contact us

Fig. 3.2 Screenshot of Privacy Policy Page Section on Twitter. Personal photo.

Approachable language seen in this section alludes to Twitter’s awareness of the issues inherent to communicating with the public. However, the goodwill this message attempts to generate is lost when the message is taken as a whole. Twitter shifts the blame onto “regulators” and absolves itself of better informative approaches by providing a deceptively informative graphic.

Looking critically at each statement reveals exceptionally low descriptive information for readers. Twitter’s existence as a public platform describes how the company operates but conveys little else, pertaining to how public platforms operate, to uninformed readers. Statements like these, which require insider knowledge to comprehend, accomplish the opposite of Twitter’s goal to empower readers. Rather, each of these statements offers an illusion of clarity. Information-seeking users must perform additional actions to find what these generalized statements mean. Selecting the hyperlinked text brings readers to specific sections of the page, at which point readers must then click additional dialogue boxes to view the information.

Addressing the section on data collection reveals additional rhetorical flourishes to persuade readers into believing Twitter's fair and informative projection. Twitter says, "You give some data, we get some data. In return we offer useful services" on their privacy policy page. Arranging the transaction in this manner alludes to some inherent knowledge and control on part of the new signee. This allusion fails when Twitter auto-enrolls new users into data collection, unless told not to specifically. New members of Twitter must actively choose to retain their data. It is more accurate to translate this sentence by saying, 'We will take your data, unless you say otherwise, and giving us this data enables us to offer more features.'

Delving further into this statement, Twitter discloses the useful services new members receive from not opting out of data collection, and these relate to user verification along with improving the services of Twitter. The generalities revealed throughout the truncated menus depict far less personally engaging experiences than a simple trade of data for services. Users supply their data and receive the possibility of receiving beneficial experiences, which contrasts with the strength of their juxtaposed one-to-one trade statement at the beginning of the section.   Opening additional dialogue menus introduce paragraph-style informational sections. These paragraphs are dense in content but only require a twelfth-grade reading level to comprehend (Readability Calculator, n.d.). Privacy policies are complex arrangements, as indicated by the length of the overall section of the website.

People new to Twitter should not have to choose between overly simplistic, vacuous statements or navigate several pages of text, various truncated dialogue boxes of no uniform length, or spend an inordinate amount of time reading to find out the details of the arrangement they auto-enroll into. The Terms of Service and cookie policy pages feature different stylistic approaches from how Twitter structures its privacy policy page. Twitter's privacy policy page features the most information for users to read and parse through, followed by the terms of service, and then cookies. The latter two pages also feature much less of an architectural webpage design. These two pages contain dense columns of text, rather than brief overviews, hypertext, and truncated menus, which makes absorbing the material more difficult.

The cookies page is exceptionally short, in comparison, but still demonstrates a lower level of effort exerted in translating cookie usage to users. The reading difficulty of this page climbed, slightly, to the collegiate level (Readability Calculator, n.d.). Terms of Service featured the most complex language to understand, requiring post-collegiate education to interpret, something the vast majority of the U.S. population does not possess. (Readability Calculator, n.d.).

Twitter's lack of transparency in the section most vital for the continuity of services for users denotes several potential avenues for consideration. Twitter may not face similar levels of public pressure to translate its use terms as it does with data privacy, given the different construction for these separate web pages. The benefits of having these essential terms explained explicitly may outweigh the detriments of abridged supplemental versions. Twitter may also prefer gatekeeping this section to avoid Terms of Service disputes with average users. In any case, the information most critical to know what are and are not permissible actions on Twitter remains within dense and difficult-to-decipher text, which contributes to poorer conceptualization of the rules of Twitter and thus leads to higher levels of MDM spread within Twitter.

Rules

The first link to Twitter’s formalized list of rules and obligations appears in a brief entry at the top of the Terms of Service page. Rather than direct users to the page, Twitter embeds the link to the page within the following disclaimer:

> If you live outside the European Union, EFTA States, or the United Kingdom, including if you live in the United States, the Twitter User Agreement comprises these Terms of Service, our Privacy Policy, the Twitter Rules and Policies, and all incorporated policies.
> If you live in the European Union, EFTA States, or the United Kingdom, the Twitter User Agreement comprises these Terms of Service, our Privacy Policy, the Twitter Rules and Policies, and all incorporated policies.

Twitter forgoes standardization of presentation in favor of users selecting the last hyperlink of their appropriate region to finally view Twitter’s rules and policies. Again, Twitter presents its audience with a summarized list of each section within the webpage. Selecting any option directs users to the appropriate section of the webpage automatically. Twitter’s policies regarding general use, platform integrity and authenticity, safety and cybercrime, and platform use guidelines are included on this page. However, the descriptive text behind these categories once again requires additional user input to select another hyperlink leading to one final webpage describing each policy’s content.

Twitter opens this section with a brief statement of purpose:

> Twitter's purpose is to serve the public conversation. Violence, harassment and other similar types of behavior discourage people from expressing themselves,

> and ultimately diminish the value of global public conversation. Our rules are to ensure all people can participate in the public conversation freely and safely.

The actual rules for tolerable actions on Twitter are brief in relation to the previous sections. Terms outlined in this section feature definitional explanations in separate, digestible paragraphs. The language utilized in these explanations, however, ranges from requiring only a high-school level reading level for their ‘Suicide or self-harm’ policy to requiring a post-graduate reading level for their ‘Perpetrators of violent attacks’ policy (Readability Calculator, n.d.). These two policies are displayed as such:

> Perpetrators of violent attacks: We will remove any accounts maintained by individual perpetrators of terrorist, violent extremist, or mass violent attacks, and may also remove Tweets disseminating manifestos or other content produced by perpetrators. Learn more.
> Suicide or self-harm: You may not promote or encourage suicide or self-harm. Learn more.

Differentiating language in the same section adds additional difficulty for average readers and contradicts their statement of purpose. Rules with uniform stylization avoid unneeded confusion, thereby better equipping readers to adequately grasp what they can and cannot do on Twitter. Twitter failing to present their rules uniformly ensures less adherence and potentially dissuades users from reading further, given that people may form assumptions about the readability of the entire T&C based on a few difficult initial sections. The public has too many barriers to satisfactorily comprehend Twitter’s rules and thus fails to serve and ensure public participation.

Rules addressing elements of MDM are sparse within Twitter’s guidelines. No text directly addresses sharing misinformative material. Twitter does have rules against synthetic and manipulated material:

> Synthetic and manipulated media: You may not deceptively share synthetic or manipulated media that are likely to cause harm. In addition, we may label

> Tweets containing synthetic and manipulated media to help people understand their authenticity and to provide additional context. Learn more.

Manipulated media likely to cause harm encompasses a broad category of material as neither of these terms is explained in this section. Selecting the provided link provides a greater explanation of these terms. However, the presentation of this information follows suit with previous shortcomings, as the material presented takes shape through several elaborate paragraphs. MDM content falls under these generalized definitions when considering fake news "is intentionally written to mislead readers to believe false information" (Shu et al., 2017). The interesting aspect to consider with synthetic and manipulated data is how this material spreads. Current research demonstrates close-knit groups represent nodes by which this information is typically spread (Dourado, 2023). Findings like this indicate a higher need to guide audiences rather than rely on top-down initiatives to inform users of potentially manipulated content, as bringing awareness of this material does not address the root issue of this material's generation and spread.

Observations

Performing this study to view how Twitter informs its userbase of its Terms and Conditions reveals stark contrasts between Twitter's stated purpose and aims to the reality of how this company interacts with those users. The ease of joining Twitter pales in comparison to the difficulty in learning how Twitter wants its users to behave. Requiring only five steps to join a platform demonstrates Twitter has the capacity to refine and streamline complicated processes yet fails to extend this refinement after users join.

Furthermore, this case study demonstrates several barriers new and older users must overcome to access the information they need to understand both permissible and impermissible behavior on Twitter. The information itself is fractured and segmented across several web pages, which is not necessary. Aggregating Twitter’s rules into a singular location would benefit average users, as the current system demonstrates inefficiencies and a lack of unified direction. Rules do not need several different webpage designs, nor do they require multiple layers of text for users to pour through to achieve a rudimentary understanding.

The difference in design and approach in Twitter's privacy policy and rules page, therefore, highlight alternative priorities akin to Twitter only serving the public insofar as to assuage widespread concern, rather than taking steps to ensure free and safe conversation. Very simply, if Twitter wanted to serve the public conversation, while diminishing discouraging behaviors, then Twitter would feature more approachable and transmissible means of educating its userbase. Instead, Twitter incorporates user agreements during the sign-up process, to effectively dissuade new users from exploring and understanding the document they agree to by completing Twitter’s account creation process. Twitter automatically enrolling new users into data harvesting and choosing less obtrusive formatting for links to their T&C document depict a company’s attempt to attain consent through exceedingly uninformative means.

Limitations

Much of this case study is constrained by the methods undertaken to critique Twitter’s T&Cs. The perspective taken in this study pertains to new members, and it is possible that Twitter’s rules and formatting are more approachable once users familiarize themselves with the platform. Clicking multiple links to multiple web pages complicates the learning process by asking users to exert more effort to find the information they seek, but older users may experience less hardship once accustomed to the website.

Observing the superficial nature of Twitter neglects alternative methods users may utilize to learn. It is possible users may acquire a better understanding of Twitter’s rules through user-to-user interactions. Simple interpersonal exchanges within the site may adequately inform users of permissible and impermissible action, so future scholarship should endeavor to understand how well average users understand the rules they agree to as well as how these users acquired their knowledge.

Device formatting remains one of the largest limitations of this study. The adoption of smart devices as opposed to desktops and laptops entails different designs for these web pages. The devices users utilize matter, such as an Android, iPhone, tablet, or other personal computer feature different layouts and methods to navigate through Twitter. I performed this study utilizing a Windows laptop, so the observations and difficulties experienced with this device may not translate perfectly to other devices. With that said, desktops and laptops provide the best viewing experience for reading documents and opening separate links. Smaller devices pose additional challenges as these devices feature less screen space and lack precise navigational tools, like computer mice and keyboards.

Reddit: r/WorldNews

Good Terms and Conditions documents are difficult to find. Plenty of T&Cs accomplish their designed purpose in defining boundaries, managing liability, and establishing itself as the primary document for either party to turn to when conflict occurs. Yet T&Cs fail to effectively translate their purpose and conditions to the vast majority of people who sign them. Instead, T&Cs occupy a detrimental space where they define designations that only lawyers readily understand.

Borrowing from the medical field, informed consent developed conceptually to ensure patients fully grasp the procedures suggested to them to avoid potential misunderstandings, given the assumption of risk patients undertake in medical procedures (Mallerdi V., 2005). When the level of potential harm passes certain thresholds, healthcare providers bear the burden to help patients reach a sufficient level of understanding regarding the outcomes of their procedure. The medical field acknowledges that people should be aware of the risks of engaging in activities where the potential for serious harm is high.

Social media platforms do not bear the same burden, despite the significant detrimental health outcomes using these online spaces can cause in users. Small changes may range from lower quality sleep, mood changes, and poorer mental health (Alonzo et al., 2021). Further investigation reveals more serious negative health outcomes as well, such as increased anxiety, depression, and suicide (Sadagheyani & Tatari, 2020). These are life-altering outcomes of participating in social media.

Though these outcomes are not endemic to everyone participating in social media, the groups displaying these traits, usually young people, deserve a higher degree of care from platform providers. U.S. society already values the importance of informed consent in the medical field, ensuring at-risk patients fully understand any and all potentially dangerous outcomes for treatments. This is done through patients and providers conversing over these potential harms before following through with such procedures. Socially normalizing the conversation surrounding informing people about the comparable risks social media use causes begins by providing information and direction.

Good T&Cs then must accomplish an additional task, finding a balance between managing liability and informing users of the harms they may encounter by participating on social media. With this in mind, this case study must analyze a user agreement that demonstrates the possibility and potential of giving users rules which they demonstrably understand, and have the capacity to follow. Adding to this, finding an analogous community to Twitter's userbase regarding topic diversity, ease of use, and the total number of users brings additional parameters to consider. The purpose of this case study is to observe how online communities act when useful guidance is both offered and transmitted well. Despite the largest social media platforms utilizing complex structures and difficult language in their T&Cs, some exceptions do exist. Reddit's satellite communities, subreddits, exist as one of such exceptions and, therefore, serve as an example ripe for analysis.

Subreddits feature varying levels of user engagement, differences in purpose, and can closely mirror the interactions seen within larger applications, like YouTube and Twitter. However, not every subreddit provides salient rules for their community. The subreddit must feature an expansive audience, permit user-generated posts and comments, have rules to follow to avoid expulsion, and demonstrate an effective path to educating new members of these rules for this community to function as a comparable example. With these criteria under consideration, r/WorldNews fulfills every point, possessing 30 million subscribed members. The forum permits users to post and comment original content. Additionally, the subreddit features various rules for participation and contains elements that promote new users to read and learn the subreddit-specific rules.

Despite no explicit contract existing for users to sign, the spirit of this endeavor is met through the social contract people abide by to avoid punishment from r/WorldNews' moderation team. General increases in subscribers indicate a willingness and interest to comply and learn how to participate in this subreddit. The following analysis concerns only elements liable to drive audience focus for the purpose of educating users about the rules of the subreddit. Elements include the visuals users see, the layout of the subreddit, and features that users interact with in navigating r/WorldNews. As this study relates to MDM reduction, only rules pertaining to user behavior within the realm of MDM reduction will qualify for critique.

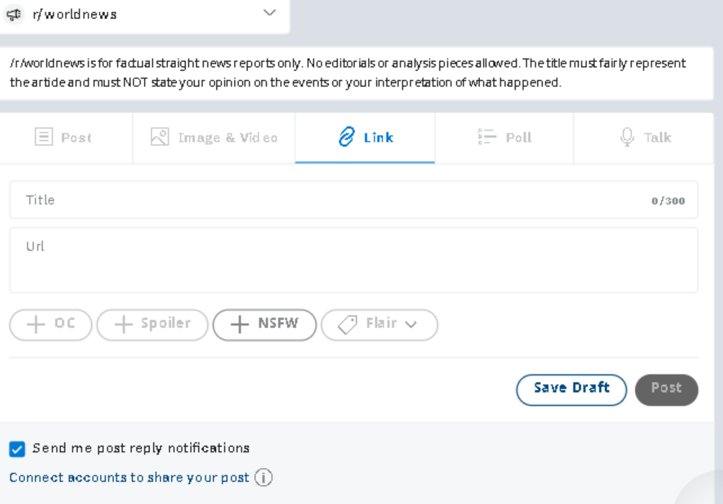
Generating posts and interacting with other users are entry points for MDM. Information users share influences topical conversations within subreddits as well as spread to larger, silent audiences often referred to as "lurkers." Lurkers are members of online communities who do not post their own content and, instead, make their presence known through Reddit's upvote and downvote features (Zhu & Dawson, 2023). Rule adoption in lurkers may take shape through tracking the popularity of high-quality posts, that abide by the rules of the subreddit. Low-quality posts then must demonstrate a lack of engagement or high levels of downvoting resulting from the community discouraging rule-breaking through lack of engagement or voting. Observing higher ratios of user comments and submissions that comply with r/WorldNews' rules then demonstrates adherence to policy and therefore indicates successful education of subreddit policies. These two realms are where r/WorldNews must effectively guide and inform its audience to reduce MDM.

Webpage Architecture

Opening the subreddit as a new user inundates the viewer with a list of postings to scroll through freely. Graphically, these posts feature few distracting elements, as the titles in posts display little more than contrasting text, black font on white background for light theme users. Excluding eye-catching elements reduces potential distractions users experience when visiting this subreddit, and r/WorldNews promotes this thematic strategy through other characteristics of the webpage. Posts exist as information cards and feature elements which include a title, the number of comments, the number of upvotes, and a direct link to the referenced article. Instead of directing people to immediately generate content, the blue banners of the sidebar draw the attention of new members from the contrasting elements compared to the visually bland format of r/WorldNews.

The sidebar functions as a brief introduction to the community. R/WorldNews captures viewer attention by reserving a few colorful elements to the sidebar, which feature sections about the community, how to filter specific topics, and rules for the subreddit. The "About Community" section provides a brief statement of purpose for the subreddit, as well as information regarding the subreddit's userbase. Subreddits, generally, feature a statement of purpose to introduce visitors to the types of content hosted on the subreddit. R/WorldNews defines its purpose as, “A place for major news from around the world, excluding US-internal news” on the homepage of the subreddit. Here, once again, the tone of the subreddit emerges. ‘Major news’ projects a different purpose as opposed to existing as a corner of the internet dedicated to daily, average events.

Interestingly enough, r/WorldNews positions a small bar below this statement to guide users to post in figure 3.3. Positioning this button before any sort of guiding text, beyond the statement of purpose, undoubtedly ensures some users attempt to post before understanding the parameters for acceptable content within this subreddit. New posters still receive direction before posting, however, as r/WorldNews includes descriptive guidance in multiple locations within the webpage users generate their posts in. Once again, the webpage’s architecture appeals to typical English-based textual directionality to push posters to read before submitting their content:



/r/worldnews is for factual straight news reports only. No editorials or analysis pieces allowed. The title must fairly represent the artide and must NOT state your opinion on the events or your interpretation of what happened.

☑ Send me post reply notifications

Connect accounts to share your post

Fig. 3.3. Screenshot of Post Creation Page on R/WorldNews. Personal photo.

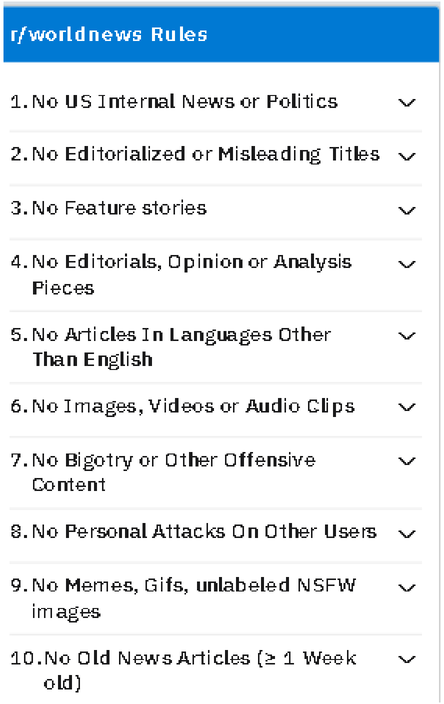
Including criteria for posts directly above the panel users interact with accomplishes one goal, providing users with guidance through encouraging members to post. Placing this text above the user submission box primes posters to, at least, glance over the rules. Though not guaranteed, intentionally placing the quintessential guiding rule of the subreddit demonstrates some level of intentional design for posters to abide by the rules, rather than ignore them all together were the text is placed further below.

Critically observing this text reveals a mix of descriptive and nuanced directions. Factual straight news, editorials, and analysis are disallowed yet not explained. However, indicating users must provide a fair representation of the article they share by withholding user opinion implies some connection to the aforementioned terms. Choosing to connect these two sentences within the same paragraph amplifies this effect. Users read this as a complete statement, rather than two separate ideas on the same topic. Furthermore, capitalizing “NOT” differentiates the word from the rest of the text, increasing audience attention due to the contrasting style.

The post submission page features additional elements simplifying adherence to r/WorldNews’ policy as well as Reddit’s site-wide policies. Sidebar changes occur when users transition to the submission webpage. Instead of featuring the normal ten rules, the list narrows to only include seven rules critical to avoiding moderative correction. Altering the sidebar to feature relevant direction promotes higher interest, as changing elements attract more attention when the norm for the subreddit is static content. Additionally, shrinking the number of requirements makes the list more approachable to posters who may lack the interest necessary to read longer blocks of text available on the homepage of the subreddit.

Rules

Comparatively, r/WorldNews introduces community rules more simplistically than Twitter in figure 3.4. The relative size and activities Twitter pursues regarding data acquisition entail higher levels of interaction between users and Twitter, so this aspect is reflected in the different T&Cs and is expected. Though subscribers to r/WorldNews engage differently, the community is lively and features a plethora of content creation and sharing. The homepage encapsulates this well, featuring several elements to promote engagement, while also offering useful sorting tools to users. The sidebar introduces new users to the purpose of r/WorldNews and remains visible whenever entering the subreddit. Additionally, this means users always have access to the rules of r/WorldNews.



r/worldnews Rules

1. No US Internal News or Politics
2. No Editorialized or Misleading Titles
3. No Feature stories
4. No Editorials, Opinion or Analysis Pieces
5. No Articles In Languages Other Than English
6. No Images, Videos or Audio Clips
7. No Bigotry or Other Offensive Content
8. No Personal Attacks On Other Users
9. No Memes, Gifs, unlabeled NSFW images
10. No Old News Articles (≥ 1 Week old)

Fig. 3.4. Screenshot of r/WorldNews' Rules List. Personal photo.

From a stylistic point, this subreddit spares users from distracting elements. None of the rules feature more than seven words in their initial presentation, which allows quick perusal to grasp the main aim of the rule. However, users may select any individual rule to activate a singular, small drop-down dialogue box. These elongated descriptions persist as well, allowing multiple long-form descriptions to remain visible at any point. The rules pertaining to reducing MDM are rules two through four and nine. Rules outside my parameters of analysis concern content cohesion and unaccepted user-to-user behavior, which do not pertain to the goals of this case study.

Rule two introduces new users to what content is banned on this subreddit and contains the largest amount of text for users to memorize, which is separated into three paragraphs:

> 2. No Editorialized or Misleading Titles
> Do not add opinion/commentary to the submission title. Don't add something that isn't covered by the article, and don't misrepresent the article. Adding a sentence from within the article that is more representative of the content is generally OK.
>
> An article's title must not be misleading. It may be removed if the title states a source's opinion as fact, or misrepresents the facts in the article or multiple other sources.
> All caps words are not allowed, except for acronyms of course.

Complex language is not present in this subreddit’s rules, requiring only high-school-level comprehension (Readability Calculator, n.d.). In addition to this, the descriptive text elaborating these rules offers useful comparisons relevant to the audience and is indicative of ground-up community building unique to subreddits. Rule three continues this theme by explaining why feature stories do not qualify as genuine sources of news within r/WorldNews:

> 3. No Feature stories
> Feature stories are journalistic reports providing more descriptive background information than a straight news report will contain. Dictionary.com: "a

> newspaper or magazine article or report of a person, event, an aspect of a major event, or the like, often having a personal slant and written in an individual style."

Rule three appeals to a common definition of Feature stories to not only filter out this type of content from the subreddit but also to inform the community of potential misinformative elements present in feature stories, such as the suggested "personal slant… [or] individual style". Introducing reasoning behind why this subreddit enforces these rules serves as a concrete example of the ability social media communities possess to connect and communicate with their userbases on a more personal and relevant level. Coupled with this, r/WorldNews demonstrates the deployment of communicative strategies to better translate meaning and reasoning to its audience, a feature other social media platforms often lack. Rule four continues to elucidate impermissible content, further refining r/WorldNews' conception of "straight news":

> 4. No Editorials, Opinion or Analysis Pieces
> /r/worldnews is for news, rather than analysis. There are several subreddits listed at the top of the page that are good for this. If the writer injects his/her opinion in the article or tries to draw any conclusion about a set of events, then it is no longer straight news and is not permitted in /r/worldnews.

Delineating between news and analysis reiterates the previous sentiments found in rules two and three. Rule four unambiguously distinguishes differences between "straight news" and analysis by drawing attention to personal effects inherent to analytical content. Reducing ambiguity through descriptive explanation benefits r/WorldNews' userbase, as it informs them of the reasons behind why certain material is banned.

Placing a rule on editorialized or misleading content second in the list suggests greater prioritization for users to understand the constraints they must follow to submit their own posts. This rule operates in direct contradiction with many features behind why MDM content spreads rapidly. R/WorldNews makes no explicit claim to combat MDM spread, yet its content rules address several common avenues misinformation spreads through. News featuring titles void of editorialization and personal opinion does not "appeal to our emotions and curiosity" nearly as effectively as how MDM content tends to (Hilary & Dumebi, 2021). People drive MDM spread, so eliminating the factors that "pique readers' interest by appealing to their emotions" dampens one aspect of what makes MDM so transmissible (Hilary & Dumebi, 2021). Though not stated explicitly in rule two, banning misleading titles addresses how MDM content "convey[s] the wrong impression if the headline is deceptive" (Hilary & Dumebi, 2021). Requiring users to adhere to rule two attacks some of the powerful aspects pertaining to the appealing nature of MDM content. Based on these findings, social media certainly possess capabilities to combat misinformative content beyond strictly excising users and content.

Analyzing in more depth, r/WorldNews demonstrates a higher propensity for showing users the differences between genuine and MDM content. Analysis of this factor considers general trends within the subreddit. Utilizing Reddit's sorting feature to display the highest upvoted posts of the month reveals adherence to r/WorldNews' rules, in addition to revealing higher engagement. People converse over content lacking much of the emotional appeal inherent to MDM content, despite the mundane nature of this subreddit's submissions.

The state of the subreddit may indicate concordance with Azzimonto & Fernandes' findings regarding threshold rules (2022). Specifically, r/WorldNews promotes unbiased content, which facilitates "agents only pay[ing] attention to sufficiently like-minded agents," which is typically seen as detrimental from an empowered consumer approach (Azzimonti & Fernandes, 2022). However, diminishing

MDM content typically spread by bots to humans through the promotion of unbiased content appears to “move opinions towards the true state” thereby reducing “both polarization and misinformation” from human agents (Azzimonti & Fernandes, 2022). Though not conclusory of the subreddit’s ability to teach, observing a strong and consistent correlation between posts that follow rules and the engagement they generate points future scholarship to potential discoveries from studying the social dynamics at play leading to these emergent situations.

Rule two’s requirement to avoid introducing personal opinions or commentary in the post titles permits greater topical clarity by removing potential personal bias factors from postings. Additionally, demanding users to interpret any articles they post sets precedent for establishing higher contextual literacy by requiring posters to read the entirety of the post they wish to submit, instead of sharing articles based on provocative titles alone. Further evidence emerges through the following sentence. Users may include text from the article they wish to share if their addition represents the article. Requiring representative text from the article itself elevates the level of comprehension to a higher degree, in the instances where users elect to do so. Instead of simply sharing information, rule two encourages users to read more critically to acquire a general understanding of their post in order to add descriptive text which accurately represents the article. Post descriptions offer another potential avenue of analysis for measuring rule adherence. Observing how many posts contain textual descriptions reveals how often posters endeavor to obtain a general understanding of the events they submit to comply with the entirety of rule two.

Analyzing rule nine requires a degree of separation not seen in the previous rules. Rules two through four offer explicit direction to guide users into compliance by reaching a similar level understand of why certain content cannot exist in r/WorldNews. Rule nine provides simple commands to follow. Instead of supplying a reason for banning memes, as included in prior rules, no explanation exists:

> 9. No Memes, Gifs, unlabeled NSFW images
> Memes may show up in image or in text form. Both are not allowed.
> Besides, porn or shock images unrelated to the story discussed will always be removed and posting them is an easy way to get banned.

Banning memes, gifs, and explicit material focuses on an avenue of MDM content not addressed through the distinct contextual explanations offered in rules two through four. Misinformative content styled as memes spread more rapidly than written content as memes require less “brain time or bandwidth” as visual images represent and communicate ideas that are “easier to digest for the viewer” (Ireland, 2018). Memes can communicate legitimate ideas to readers, and have the capacity to relay major news events. It is then interesting to see a hardline stance against posting memes within the subreddit. As no reason is offered, r/WorldNews may understand the pernicious nature of memetic content and its unique ability to spread MDM content far more effectively than the traditional posting style seen in this subreddit. Other potentialities exist though, as r/WorldNews may also desire to foster more serious and dedicated audience members typically not seen in other meme oriented communities, such as Twitter. Regardless of intentionality, banning memes eliminates a powerful tool MDM spreaders typically use on vulnerable populations.

Rule ten on its own does not appear to reduce MDM content, yet does point subscribers and lurkers to observe current news, which shields against one unique element of MDM memes:

> 10. No Old News Articles (≥ 1 Week old)
> Old News is any news older than one week.

Though relatively short, rule ten follows rule nine in design, as no explanation offers a descriptive explanation to users for following this directive. Persistent major news, such as the posts on the Russian—Ukraine conflict, indicate r/WorldNews does permit ongoing stories to remain. The distinction then must pertain to stories that are periodic and episodic. This subreddit supports addressing stories that span periods of time when those stories develop daily, yet it appears to not oppose episodic developments which emerge briefly and fade thereafter, seen with foreign-state declarations.

When viewed concordantly rules nine and ten provide unique protection against MDM memes. Specifically, malicious actors utilize memes to generate "media spectacle [which] is spread and perpetuated by networked communities" (Mihailidis & Viotty, 2017). The key point r/WorldNews' rules address is the perpetual nature of memes and tropes which emerge through evolving memetic content. A clear line emerges with this understanding. Many scholars voice concern for improving media literacy within social media (Connaway et al., 2017; Carmi et al., 2020; Apuke et al., 2022). However, little consensus exists for determining what level of digital literacy internet communities must attain or what average people can acquire to achieve successful communal resistance from MDM memes.

Observing the stark differences between rules two through four and rules nine and ten indicate the potential for determining what average people can reasonably learn and follow. Direction within this community demonstrates average posters possess enough capacity to understand and follow rules concerning non-emotionally evocative content leading to successful communal engagement. However, content liable to carry more emotional impact is simply banned without explanation. The differences in how these rules are described indicate people generally lack the ability to parse true and false information delivered through memes, even if rules on memes bore similar descriptors to rules two through four.

Observations

My analysis sheds light on the potential that communal rules hold to inform audiences. Rules are not just text to glance over online. Spaces exist demonstrating worthy observable communal interaction which highlights the efficacy of informing participants on how to be good members within a given community. This should not surprise many, given offline society is not inherently chaotic and the vast majority of people follow the law and enjoy the benefits of doing so. Online communications introduce new potentialities for interaction, but the rapidity with which information is transferred online is not a new element.

Consider the place of automobiles in revolutionizing industry and civilian movement. Chaos is not inherent to rapid communication but can highlight the inadequacies of current systems we participate in, such as how roads required advancement and necessitated traffic laws to operate smoothly. Similar insights emerge online if the same lens is used. Reddit’s r/WorldNews demonstrates online communities can succeed even with restrictive content policies. People can learn to cohesively interact together online as well as we do offline. The evocative nature of memes leads people to respond more emotionally and are therefore banned from r/WorldNews. Similarly, road rage, break checking, and other inciting behaviors are unlawful. We recognize, as demonstrated through r/WorldNews, that emotionally charged communication leads to worse communal outcomes and this finding is inherent to online and offline society.

Educating people, then, follows as one effective tool for improving online communications by facilitating healthier interaction and should bear more focus from the largest platforms than they currently do. Online toxicity, manipulation, and bullying have existed in online communities for nearly three decades. Despite this, online platforms do little to reduce this occurrence apart from removing people from the platform entirely when they could do more to educate their userbase.

Limitations

This case study examined the architectural features of r/WorldNews to understand possible subreddit features leading to healthier communication online. As such, my findings may not translate to other fields, as the levels of audience cultivation pooling from Reddit at large and then aggregating further to r/WorldNews' smaller community may not represent average people as well as other social media spaces. Additionally, communication and rule adherence in this community may not be static. Different societal events may influence how well people adhere to the rules of this subreddit.

Issues may also emerge from moderation. My case study of r/WorldNews does not encapsulate how moderation influences communal cohesion. Subreddits feature involved human overview of content created on these sites. Auto-moderation from content filters by bots must also influence how the subreddit functions. Future work into how these moderating forces interact with human users will elucidate how well the subreddit’s presentation and formation of rules translates to its userbase. Both posts and comments require analysis to conclusively determine the efficacy of communal rule on how well online communication between users can improve.

CHAPTER FOUR

Assessing Cases and Future Direction

Conclusion

Concluding this project requires a final comparison of both sections to evaluate what my findings provide for future scholarship in this area. Understanding how IBR and T&C documents influence MDM trends is limited by the lack of previous literature surrounding IBR and T&C documents. My contributions to this field elucidate some potential avenues to pursue in future studies of MDM content reduction strategies. Harm reduction is a prime driving force behind both topics I included in this project. In pursuit of that goal, I attempted to reveal some elements known to reduce MDM spread that these documents have the unique ability to promote and enhance. Performing these case studies provides valuable insight into potential large-scale interventions of misinformative material.

Current trends in MDM research indicate inoculating populations from its influence is most effectively executed with top-down content interventions. Though these interventions, such as deplatforming, do not solve the root issue, they do tend to limit the severity of the problem. Notoriety and communal penetration are key factors to determine how harmful a piece of MDM content is. Reducing either facet limits the impact of misinformation and lowers the overall harm caused by these types of information. As research shows, average people are not equipped with the knowledge or strategies to resist misinformation, and cannot reasonably be expected to do so. Campaigns promoting greater knowledge and understanding do not, currently, spread deeply enough into U.S. society to remedy this problem.

Americans deserve greater affordances and an expansion of traditional protections and rights to correct the violations occurring daily on the Internet. As such, we must rely on the government to intervene when the public lacks the capacity to tackle pervasive, systemic issues. The insurrectionist invasion of the U.S. Capitol on January 6th highlights the necessity to update our strategies in combating propaganda and MDM content at large. We cannot tolerate actors deluding and directing large segments of the population to overthrow the systems we rely on to function. Governments aim to protect civilians from malicious actors offline, and similar protections must bridge to online communities as well. Harms online increasingly influences offline interactions, so it behooves researchers to study this avenue of intervention to offer guidance for future policy decisions.

My case studies on Internet Bills of Rights revealed a stark advancement and attention to address the emergent threats we now face in online communities. To address my first research question, the EU’s IBR does provide meaningful ways for people to control their data, effectively fulfilling its purpose. In terms of data control, there are two main lines to distinguish. Adults and children represent two groups who will always deserve separate treatment. This IBR recognizes this through its inclusions to prioritize the safety of children. Enshrining the necessity for age-appropriate specs for children to occupy provides a useful buffer to content that would otherwise harm children, while still allowing them to participate in online environments. Coupling this directive with firm limits to avoid tracking and possible algorithmically manipulating the content they see also benefits children, as they are then allowed to grow and learn without potentially problematic influence of algorithms.

Special inclusions aside, general changes in this bill also benefit adults and children alike. Demanding transparency of social media algorithms pertaining to how they function as well as when users interact with them, provides users with insight into why they see the content they do. Utilizing the broad theme of the bill, this inclusion likely aims to encourage greater understanding into the inner workings of the internet. Having more decision-making power to shape one's internet experience encourages a truer diverse body of voices on the internet to emerge. Currently, algorithms shape much of the conversations, as promoted content utilizes engagement to decide which stories to show to people. Avoiding this filtering process to instead feature more democratic means of engagement also aids in users having more control over their experience on the internet, their data they wish to share, and the information they gather from others.

My second question features more barriers to answer, as the U.S. features a different style of governance relating to private social media entities. Demanding social media companies to put more effort into removing MDM content represents an exertion of power that is not legally possible. Therefore, this aspect of the IBR cannot currently translate over. Additionally, banning unlawful and pervasive tracking again requires more governmental intrusion into private entities that the U.S. does not permit. Should a solution emerge that does not violate this principle, then the inclusions within this bill certainly aid many of the communication issues present in the U.S. Protecting children from tracking provides similar benefits already listed for the EU. Protecting communication data will make phishing and data theft attempts more difficult to execute successfully. Featuring some level of involvement with social media companies to reduce the amount of influence MDM has on these spaces is likely not possible without investment from governmental resources.

Denying any sort of protection the government may offer from a publicly occupied space, offline or online, contrasts with the U.S. government's duty to protect its civilians from threats domestic and abroad. The main contention of this thesis emerges once again, at what point do the harms become intolerable and necessitate intervention? Insurrectionist riots occurring every election cycle is a dismal thought to entertain, yet if no strategies are formulated or changes to online communication occur, the status quo will remain the same.

My last research question demands much less analysis, given the existence of useful, transmissible, and educational Terms and Conditions agreements already exist. The fact that most T&C's feature poor communicative tools in presentation alludes to the simple answer that they need not inform users for the entities drafting these documents to succeed as businesses. Rather, their complexity, in the case of Twitter, appears to take advantage of the public's inability to parse through the information Twitter presents. Inclusions, like automatically enrolling in data harvesting, point to a more malicious-driven nature of the interaction. Improving T&C's, therefore, is simple, present rules to people before allowing them to participate in your website. Research already shows significant personal harm can occur from interacting in these spaces, so mandates could utilize this avenue for mandating a more informative and person-oriented induction experience. Some barriers do emerge though, as children cannot legally offer consent in many of these cases, though some permit parental approval as appropriate means of consenting. Authenticating the identity of an anonymous internet user is not in the purview of social media companies to fulfill.

Additionally, less intrusive means of updating T&C documents can be as simple as providing brief and explicit summaries of the duties and obligations both parties must adhere to. Using simple language, informative videos, or even making the rules a dedicated element in each page users visit ensures users are at the very least exposed to proper guidelines. This would be an improvement over the current state of Twitter's rules webpage, as demonstrated in chapter whatever. These changes are simple, and if this case study serves any evidentiary function, then it points to the potentiality to demand civilians are properly informed and consent to the documents they sign.

My case studies present two points of improvement, regarding the digital rights of civilians. Underlying both, however, is the necessity of these changes to combat the real and present danger MDM content represents. Current solutions to dealing with MDM actors have failed to correct the issue. MDM content still infests many parts of online communication, be it in the communities themselves or existing as separate resources unknowing users visit. No governing body exists to point out the falsehoods present on the internet, and I argue this situation demands intervention.

No agency can remove all MDM content, but previous interventions into other communal interests tend to benefit the public. The FDA, EPA, DMV, and others all broadly represent the public's interest and perform net-benefits to society. We recognize the importance of having safe consumables, minimizing environmental damage, and having some base level of education to operate a vehicle while driving. Why do we then neglect the importance of a proper education into using the Internet to avoid the clear and present harms abundant online?

To conclude, this thesis sought to compare separate instances of possible avenues to improving how people interact with the internet. Comparing IBR's represents a top-down solution to many of the issues of online communication and communal trends. As the EU's new IBR is the first of its kind to address data and MDM issues online, further study over its lifetime will provide extremely valuable insight into limiting the impacts of misinformation. Comparing Terms and Conditions agreements represents an empowered-individualistic perspective to better involve and equip the public with a higher degree of digital literacy to resist MDM content. Both avenues represent useful tools to refine through further study in crafting effective solutions to protect societies from the divisive harms inflicted through misinformation. As such, both Internet Bills of Rights and Terms and Conditions documents deserve greater study than they currently receive.

BIBLIOGRAPHY

Agarwal, S., Ananthakrishnan, U. M., & Tucker, C. E. (2022). *Deplatforming and the Control of Misinformation: Evidence from Parler* (SSRN Scholarly Paper No. 4232871). https://doi.org/10.2139/ssrn.4232871

Ali, A., & Qazi, I. A. (2022). Digital Literacy and Vulnerability to Misinformation: Evidence from Facebook Users in Pakistan. *Journal of Quantitative Description: Digital Media*, *2*. https://doi.org/10.51685/jqd.2022.025

Allison, K., & Bussey, K. (2020). Communal Quirks and Circlejerks: A Taxonomy of Processes Contributing to Insularity in Online Communities. *Proceedings of the International AAAI Conference on Web and Social Media*, *14*, 12–23. https://doi.org/10.1609/icwsm.v14i1.7275

Alonzo, R., Hussain, J., Stranges, S., & Anderson, K. K. (2021). Interplay between social media use, sleep quality, and mental health in youth: A systematic review. *Sleep Medicine Reviews*, *56*, 101414. https://doi.org/10.1016/j.smrv.2020.101414

Al-Zahrani, A. (2015). Toward Digital Citizenship: Examining Factors Affecting Participation and Involvement in the Internet Society among Higher Education Students. *International Education Studies*, *8*(12), 203–217.

Anwar, A., Kee, D. M. H., & Ijaz, M. F. (2022). Social Media Bullying in the Workplace and Its Impact on Work Engagement: A Case of Psychological Well-Being. *Information*, *13*(4), Article 4. https://doi.org/10.3390/info13040165

Apuke, O. D., Omar, B., & Asude Tunca, E. (2022). Literacy Concepts as an Intervention Strategy for Improving Fake News Knowledge, Detection Skills, and Curtailing the Tendency to Share Fake News in Nigeria. *Child & Youth Services*, *0*(0), 1–16. https://doi.org/10.1080/0145935X.2021.2024758

Azzimonti, M., & Fernandes, M. (2022). Social media networks, fake news, and polarization. *European Journal of Political Economy*, 102256. https://doi.org/10.1016/j.ejpoleco.2022.102256

Bell, V. (2007). Online information, extreme communities and internet therapy: Is the internet good for our mental health? *Journal of Mental Health*, *16*(4), 445–457. https://doi.org/10.1080/09638230701482378

Bovet, A., & Makse, H. A. (2019). Influence of fake news in Twitter during the 2016 US presidential election. *Nature Communications*, *10*(1), Article 1. https://doi.org/10.1038/s41467-018-07761-2

Braghieri, L., Levy, R., & Makarin, A. (2022). Social Media and Mental Health. *American Economic Review*, *112*(11), 3660–3693. https://doi.org/10.1257/aer.20211218

Buiten, M. C. (2022). *Combating disinformation and ensuring diversity on online platforms: Goals and limits of EU platform regulation*.

Carmi, E., Yates, S. J., Lockley, E., & Pawluczuk, A. (2020). Data citizenship: Rethinking data literacy in the age of disinformation, misinformation, and malinformation. *Internet Policy Review*, *9*(2), 1–22. https://doi.org/10.14763/2020.2.1481

Celeste, E. (2019). Digital constitutionalism: A new systematic theorisation. *International Review of Law, Computers & Technology*, *33*(1), 76–99. https://doi.org/10.1080/13600869.2019.1562604

Celliers, M., & Hattingh, M. (2020). A Systematic Review on Fake News Themes Reported in Literature. In M. Hattingh, M. Matthee, H. Smuts, I. Pappas, Y. K. Dwivedi, & M. Mäntymäki (Eds.), *Responsible Design, Implementation and Use of Information and Communication Technology* (pp. 223–234). Springer International Publishing. https://doi.org/10.1007/978-3-030-45002-1_19

Cinelli, M., Pelicon, A., Mozetič, I., Quattrociocchi, W., Novak, P. K., & Zollo, F. (2021). Dynamics of online hate and misinformation. *Scientific Reports*, *11*(1), Article 1. https://doi.org/10.1038/s41598-021-01487-w

Connaway, L. S., Julien, H., Seadle, M., & Kasprak, A. (2017). Digital literacy in the era of fake news: Key roles for information professionals. *Proceedings of the Association for Information Science and Technology*, *54*(1), 554–555. https://doi.org/10.1002/pra2.2017.14505401070

Creswick, H., Dowthwaite, L., Koene, A., Perez Vallejos, E., Portillo, V., Cano, M., & Woodard, C. (2019). "… They don't really listen to people": Young people's concerns and recommendations for improving online experiences. *Journal of Information, Communication and Ethics in Society*, *17*(2), 167–182. https://doi.org/10.1108/JICES-11-2018-0090

Critchlow, N., Moodie, C., Stead, M., Morgan, A., Newall, P. W. S., & Dobbie, F. (2020). Visibility of age restriction warnings, harm reduction messages and terms and conditions: A content analysis of paid-for gambling advertising in the United Kingdom. *Public Health*, *184*, 79–88. https://doi.org/10.1016/j.puhe.2020.04.004

Daimi, K., & Peoples, C. (Eds.). (2021). *Advances in Cybersecurity Management*. Springer International Publishing. https://doi.org/10.1007/978-3-030-71381-2

Dori-Hacohen, S., Sung, K., Chou, J., & Lustig-Gonzalez, J. (2021). Restoring Healthy Online Discourse by Detecting and Reducing Controversy, Misinformation, and Toxicity Online. *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2627–2628. https://doi.org/10.1145/3404835.3464926

Dourado, T. (2023). Who Posts Fake News? Authentic and Inauthentic Spreaders of Fabricated News on Facebook and Twitter. *Journalism Practice*, *0*(0), 1–20. https://doi.org/10.1080/17512786.2023.2176352

Dumitru, E.-A. (2020). Testing Children and Adolescents' Ability to Identify Fake News: A Combined Design of Quasi-Experiment and Group Discussions. *Societies*, *10*(3), Article 3. https://doi.org/10.3390/soc10030071

European Declaration on Digital Rights and Principles | Shaping Europe's digital future. (n.d.). Retrieved January 15, 2023, from https://digital-strategy.ec.europa.eu/en/library/european-declaration-digital-rights-and-principles

Finckenstein, K. von, & Menzies, P. (2022). *Social media responsibility and free speech: A new approach for dealing with 'Internet harms'* (Canada) [Report]. Macdonald-Laurier Institute. https://apo.org.au/node/318140

Gradoń, K. T., Hołyst, J. A., Moy, W. R., Sienkiewicz, J., & Suchecki, K. (2021). Countering misinformation: A multidisciplinary approach. *Big Data & Society*, *8*(1), 205395172110138. https://doi.org/10.1177/20539517211013848

Güera, D., & Delp, E. J. (2018). Deepfake Video Detection Using Recurrent Neural Networks. *2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 1–6. https://doi.org/10.1109/AVSS.2018.8639163

Hajli, N., Saeed, U., Tajvidi, M., & Shirazi, F. (2022). Social Bots and the Spread of Disinformation in Social Media: The Challenges of Artificial Intelligence. *British Journal of Management*, *33*(3), 1238–1253. https://doi.org/10.1111/1467-8551.12554

Hilary, I. O., & Dumebi, O.-O. (2021). Social Media as a Tool for Misinformation and Disinformation Management. *Linguistics and Culture Review*, *5*(S1), Article S1. https://doi.org/10.21744/lingcure.v5nS1.1435

Howard, P. N., Neudert, L.-M., Prakash, N., & Vosloo, S. (n.d.). *Digital misinformation / disinformation and children*.

Huang, B., & Carley, K. M. (2020). *Disinformation and Misinformation on Twitter during the Novel Coronavirus Outbreak* (arXiv:2006.04278). arXiv. https://doi.org/10.48550/arXiv.2006.04278

Informing Effective Messaging: Rural School Personnel's Attitudes, Beliefs, &amp; Influences Related to COVID-19 Vaccination - ProQuest. (n.d.). Retrieved February 24, 2023, from https://www.proquest.com/openview/2f4c110dcf442a8e488f31defc5e28f7/1?pq-origsite=gscholar&cbl=18750&diss=y

Iqbal, W., Tyson, G., & Castro, I. (2022). Looking on Efficiency of Content Moderation Systems from the Lens of Reddit's Content Moderation Experience During COVID-19. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4007864

Ireland, S. (2018). Fake news alerts: Teaching news literacy skills in a meme world. *The Reference Librarian*, *59*(3), 122–128. https://doi.org/10.1080/02763877.2018.1463890

Jhaver, S., Boylston, C., Yang, D., & Bruckman, A. (2021). Evaluating the Effectiveness of Deplatforming as a Moderation Strategy on Twitter. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW2), 381:1-381:30. https://doi.org/10.1145/3479525

Katz, J. E., & Rice, R. E. (2002). *Social Consequences of Internet Use: Access, Involvement, and Interaction*. MIT Press.

Khan, M. L., & Idris, I. K. (2019). Recognise misinformation and verify before sharing: A reasoned action and information literacy perspective. *Behaviour & Information Technology*, *38*(12), 1194–1212. https://doi.org/10.1080/0144929X.2019.1578828

Kirchengast, T. (2020). Deepfakes and image manipulation: Criminalisation and control. *Information & Communications Technology Law*, *29*(3), 308–323. https://doi.org/10.1080/13600834.2020.1794615

Kopp, D. C. (n.d.). *Submission to the Electoral Matters Committee of the Parliament of Victoria*. 14.

Kraut, R., Patterson, M., Lundmark, V., Kiesler, S., Mukophadhyay, T., & Scherlis, W. (1998). Internet paradox: A social technology that reduces social involvement and psychological well-being? *American Psychologist*, *53*, 1017–1031. https://doi.org/10.1037/0003-066X.53.9.1017

Larson, H. J. (2018). The biggest pandemic risk? Viral misinformation. *Nature*, *562*(7727), 309–309. https://doi.org/10.1038/d41586-018-07034-4

Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., & Zittrain, J. L. (2018). The science of fake news. *Science*, *359*(6380), 1094–1096. https://doi.org/10.1126/science.aao2998

Livingstone, S. (2013). Online risk, harm and vulnerability: Reflections on the evidence base for child Internet safety policy. *ZER: Journal of Communication Studies*, *18*(35), Article 35.

López-Silva, P., & Valera, L. (2022). *Protecting the Mind: Challenges in Law, Neuroprotection, and Neurorights*. Springer Nature.

Lovato, J., Lopez, G. S., Rogers, S. P., Haq, I. U., & Onaolapo, J. (n.d.). *Diverse Misinformation: Impacts of Human Biases on Detection of Deepfakes on Networks*.

Luger, E., Moran, S., & Rodden, T. (2013). Consent for all: Revealing the hidden complexity of terms and conditions. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2687–2696. https://doi.org/10.1145/2470654.2481371

Mallardi, V. (2005). [The origin of informed consent. *Acta otorhinolaryngologica Italica*, *25*(5), 312–327.

Mapping global trends in vaccine confidence and investigating barriers to vaccine uptake: A large-scale retrospective temporal modelling study | Elsevier Enhanced Reader. (n.d.). https://doi.org/10.1016/S0140-6736(20)31558-0

Marco Civil English Version. (n.d.). Public Knowledge. Retrieved January 23, 2023, from https://publicknowledge.org/policy/marco-civil-english-version/

Martins dos Santos, B. (2021). *An assessment of the role of Marco Civil's Intermediary Liability Regime for the Development of the Internet in Brazil* (SSRN Scholarly Paper No. 4023824). https://papers.ssrn.com/abstract=4023824

Mihailidis, P., & Viotty, S. (2017). Spreadable Spectacle in Digital Culture: Civic Expression, Fake News, and the Role of Media Literacies in "Post-Fact" Society. *American Behavioral Scientist*, *61*(4), 441–454. https://doi.org/10.1177/0002764217701217

Myers, N. (2021). Information Sharing and Community Resilience: Toward a Whole Community Approach to Surveillance and Combatting the "Infodemic." *World Medical & Health Policy*, *13*(3), 581–592. https://doi.org/10.1002/wmh3.428

Nikolov, D., Flammini, A., & Menczer, F. (2021). Right and left, partisanship predicts (asymmetric) vulnerability to misinformation. *Harvard Kennedy School Misinformation Review*. https://doi.org/10.37016/mr-2020-55

Pagoto, S., Waring, M. E., & Xu, R. (2019). A Call for a Public Health Agenda for Social Media Research. *Journal of Medical Internet Research, 21*(12), e16661. https://doi.org/10.2196/16661

Parris, L., Lannin, D. G., Hynes, K., & Yazedjian, A. (2022). Exploring Social Media Rumination: Associations With Bullying, Cyberbullying, and Distress. *Journal of Interpersonal Violence, 37*(5–6), NP3041–NP3061. https://doi.org/10.1177/0886260520946826

Pascual-Ferrá, P., Alperstein, N., Barnett, D. J., & Rimal, R. N. (2021). Toxicity and verbal aggression on social media: Polarized discourse on wearing face masks during the COVID-19 pandemic. *Big Data & Society*, *8*(1), 20539517211023532. https://doi.org/10.1177/20539517211023533

Pennycook, G., & Rand, D. G. (2022). Accuracy prompts are a replicable and generalizable approach for reducing the spread of misinformation. *Nature Communications*, *13*(1), Article 1. https://doi.org/10.1038/s41467-022-30073-5

Primack, B. A., Karim, S. A., Shensa, A., Bowman, N., Knight, J., & Sidani, J. E. (2019). Positive and Negative Experiences on Social Media and Perceived Social Isolation. *American Journal of Health Promotion, 33*(6), 859–868. https://doi.org/10.1177/0890117118824196

Readability Calculator. (n.d.). Retrieved March 10, 2023, from https://www.wordcalc.com/readability/

Reddy, P., Sharma, B., & Chaudhary, K. (2020). Digital Literacy: A Review of Literature. *International Journal of Technoethics (IJT)*, *11*(2), 65–94. https://doi.org/10.4018/IJT.20200701.oa1

Redeker, D., Gill, L., & Gasser, U. (2018). Towards digital constitutionalism? Mapping attempts to craft an Internet Bill of Rights. *International Communication Gazette*, *80*(4), 302–319. https://doi.org/10.1177/1748048518757121

Sadagheyani, H. E., & Tatari, F. (2020). Investigating the role of social media on mental health. *Mental Health and Social Inclusion*, *25*(1), 41–51. https://doi.org/10.1108/MHSI-06-2020-0039

Schneble, C. O., Favaretto, M., Elger, B. S., & Shaw, D. M. (2021). Social Media Terms and Conditions and Informed Consent From Children: Ethical Analysis. *JMIR Pediatrics and Parenting, 4*(2), e22281. https://doi.org/10.2196/22281

Schreiber, A. (2022). Civil Rights Framework of the Internet (BCRFI; Marco Civil da Internet): Advance or Setback? Civil Liability for Damage Derived from Content Generated by Third Party. In M. Albers & I. W. Sarlet (Eds.), *Personality and Data Protection Rights on the Internet: Brazilian and German Approaches* (pp. 241–266). Springer International Publishing. https://doi.org/10.1007/978-3-030-90331-2_10

Shaw, C. R. (n.d.). *Decentralized Social Networks: Pros and Cons of the Mastodon Platform*. 6.

Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake News Detection on Social Media: A Data Mining Perspective. *ACM SIGKDD Explorations Newsletter*, *19*(1), 22–36. https://doi.org/10.1145/3137597.3137600

Stein, J., Keuschnigg, M., & van de Rijt, A. (2023). Network segregation and the propagation of misinformation. *Scientific Reports*, *13*(1), Article 1. https://doi.org/10.1038/s41598-022-26913-5

Steinfeld, N. (2016). *"I agree to the terms and conditions": (How) do users read privacy policies online? An eye-tracking experiment. Computers in Human Behavior, 55, 992–1000. https://doi.org/10.1016/j.chb.2015.09.038*

Tesfay, W. B., Hofmann, P., Nakamura, T., Kiyomoto, S., & Serna, J. (2018). PrivacyGuide: Towards an Implementation of the EU GDPR on Internet Privacy Policy Evaluation. *Proceedings of the Fourth ACM International Workshop on Security and Privacy Analytics*, 15–21. https://doi.org/10.1145/3180445.3180447

Twenge, J. M., Haidt, J., Lozano, J., & Cummins, K. M. (2022). Specification curve analysis shows that social media use is linked to poor mental health, especially among girls. *Acta Psychologica*, *224*, 103512. https://doi.org/10.1016/j.actpsy.2022.103512

Tyler, T., & De Cremer, D. (2009). Ethics and rule adherence in groups. In *Psychological perspectives on ethical behavior and decision making* (pp. 215–232). Information Age Publishing, Inc.

United States: Number of Facebook users 2027. (n.d.). Statista. Retrieved November 9, 2022, from https://www.statista.com/statistics/408971/number-of-us-facebook-users/

Uyheng, J., & Carley, K. M. (2021). Characterizing network dynamics of online hate communities around the COVID-19 pandemic. *Applied Network Science*, *6*(1), 20. https://doi.org/10.1007/s41109-021-00362-x

Vasu, N., Ang, B., & Jayakumar, S. (2018). *Drums: Distortions, Rumours, Untruths, Misinformation, And Smears*. World Scientific.

Wei, L., Gong, J., Xu, J., Eeza Zainal Abidin, N., & Destiny Apuke, O. (2023). Do social media literacy skills help in combating fake news spread? Modelling the moderating role of social media literacy skills in the relationship between rational choice factors and fake news sharing behaviour. *Telematics and Informatics*, *76*, 101910. https://doi.org/10.1016/j.tele.2022.101910

Wu, W., Lyu, H., & Luo, J. (2021). Characterizing Discourse about COVID-19 Vaccines: A Reddit Version of the Pandemic Story. *Health Data Science*, *2021*, 2021/9837856. https://doi.org/10.34133/2021/9837856

Yoo, C. S. (2009). Free Speech and the Myth of the Internet as an Unintermediated Experience. *George Washington Law Review*, *78*(4), 697–773.

Zhao, R., Zhou, A., & Mao, K. (2016). Automatic detection of cyberbullying on social networks based on bullying features. *Proceedings of the 17th International Conference on Distributed Computing and Networking*, 1–6. https://doi.org/10.1145/2833312.2849567

Zhu, J., & Dawson, K. (n.d.). Lurkers versus posters: Perceptions of learning in informal social media-based communities. *British Journal of Educational Technology*, *n/a*(n/a). https://doi.org/10.1111/bjet.13303