



Social Media - friend or foe?

Mutale Mwila

A dissertation submitted to the Faculty of Commerce, University of Cape Town, in partial fulfilment of the requirements for the degree of Master of Philosophy specialising in Financial Technology.

Supervised by Associate Professor Co-Pierre Georg

School of Economics, Faculty of Commerce

University of Cape Town

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

Declaration

I, Mutale Mwila, declare that this thesis is composed of my original work; it contains no material previously published or written by another person except where the text has made due reference. This thesis is submitted to the University of Cape Town for the Master of Philosophy specialising in Financial Technology degree.

Signed by candidate

23/09/2023

Signature

Date

Abstract

The use of social media has become a part of our daily lives. Thus, there is growing concern over social media platforms' impact on individuals and society. Beginning with an introductory chapter outlining research objectives, subsequent sections delve into various facets: Chapter two delineates the methodology and theoretical framework applied in this thesis. Chapter three scrutinises how the architecture of social media platforms exploits human vulnerabilities, leading to pervasive societal and individual harm. Chapter four explores social media's influence on collective well-being, encompassing its role in elections, global conflict, and information dissemination through major digital platforms. Chapter five investigates the psychological effects of social media on individuals, focusing on diverse literature and its impact on children's development. Moreover, this thesis evaluates proposed regulatory measures for governing social media platforms, assessing their practicality. Concluding remarks address limitations and suggest avenues for future research. By synthesising a wide array of perspectives, this thesis offers a comprehensive exploration of social media's implications, shedding light on its profound effects on societal dynamics and individual well-being.

Acknowledgements

I want to thank the following individuals without whose support this thesis could never have happened.

I am grateful to my supervisor, Associate Professor Co-Pierre Georg, for the guidance and insight you provided throughout this project and degree. Thank you for instilling in your students the desire to make the world a better place.

To my parents and siblings, my biggest cheerleaders, without your love and support, I would never have been able to reach this point in my studies; I am eternally grateful.

To Timothy, thank you for holding my hand throughout this journey. We'll see what's about to happen next.

Elsie, Njavwa and Katai. Thank you for being my board of counsel and for the laughter and encouragement.

And to the rest of my family and friends, thank you for all the endless support you provided throughout my studies.

Table of Contents

Social Media - friend or foe?	1
Declaration.....	i
Abstract.....	ii
Acknowledgements.....	iii
Table of Contents.....	iv
Chapter 1: Introduction.....	1
Chapter 2: Method	5
Chapter 3: The Rise of Social Media.....	13
Chapter 4: Societal Implications of Social Media	20
Chapter 5: Impact of Social Media on Individual Well-being.....	38
Chapter 6: Regulation	45
Chapter 7: Limitations and Future Research	56
Chapter 8: Conclusion.....	57
References.....	59

Chapter 1: Introduction

Social media has weaved itself into the fabric of society, dramatically changing how humans communicate and entertain themselves. Today, 4.62 billion people, more than half of the world's population, use social media (Kemp, 2022). Platforms such as Twitter, Facebook and YouTube have profoundly altered the social, political, and informational practices of individuals and institutions across the globe (Vallor, 2012). This ubiquitous use of social media prompts a greater understanding of these platforms' impact. While most social media companies tout their ability to connect people and ideas worldwide, a darker side has emerged, evidenced by multiple events. From the rampant spread of misinformation during the COVID-19 pandemic to the genocide in Myanmar, the adverse effects of social media platforms are becoming increasingly apparent. Society is only just beginning to recognise the implications of the pervasiveness of social media use and what it means for personal safety and collective well-being.

This thesis provides a comprehensive exploration of the ramifications of social media. It begins with an introductory chapter outlining the background and research objectives. The methodology and theoretical framework utilised in this research are detailed in the second chapter. The subsequent chapters delve into specific aspects:

- Chapter three scrutinises how the architecture of social media platforms exploits human vulnerabilities for financial gain, shedding light on their design flaws contributing to societal and individual harm.
- Chapter four examines social media's impact on collective well-being, including its influence on elections, global conflict, and information dissemination via major tech companies.
- Chapter five focuses on social media's effects on individuals' psychological well-being, encompassing diverse literature on its influence and its effects on children's development.
- Furthermore, the thesis evaluates proposed regulatory measures for overseeing social media platforms, assessing their practicality and real-world applicability. Finally, it concludes by outlining the limitations of the thesis and suggesting avenues for future research.

As this thesis demonstrates, the challenges associated with social media platforms are multi-layered, encompassing ethics, policy, and business. The solution to such challenges lies at the intersection of ethical technology practices, an understanding of human behaviour, and radical systems reform.

Background

Social media use is a global phenomenon, and as internet access around the world increases, so will the number of social media users. No technology since television has dramatically reshaped how people get information and spend their time (Allcott et al., 2020). The ubiquitous use of social media has provoked both optimism about potential societal benefits and concern about the harms created by these digital platforms. It is therefore essential to understand what social media is and what common characteristics social media platforms have.

Social media refers to websites and internet-based applications built on the technological foundations of Web 2.0. Social media platforms allow for creating and exchanging user-generated content and user participation in social networking (Kaplan and Haenlein, 2009).

The social network is the typical form of social media (Meshi et al., 2015). Social network platforms are web-based services that allow individuals to create public profiles and connect with other users (Boyd and Ellison, 2007). The information included on a profile varies by social network. However, it commonly consists of a profile picture, username, and information about demographics and personality, such as gender, education, employment, and interests. Social networks can have specific uses. For instance, LinkedIn centres careers, Twitter has a microblogging focus, and Instagram is used for photo sharing.

Other major categories of social media include media sharing platforms, social news, and collaborative content. Media sharing platforms, such as YouTube, provide a platform for users to share media such as music and videos. This category often overlaps with social networks because media-sharing platforms allow for profiles, comments, or feedback on posted content. For instance, because it is a social network focused on photo sharing, Instagram is considered a media-sharing platform. Social news platforms, such as Reddit, allow users to discuss news. Lastly, collaborative content platforms like Wikipedia collectively enable users to contribute to content on a public website. Users can generate, edit, and delete content but do not interact socially as on other platforms (Allcott et al., 2020).

Social media platforms have common characteristics, such as low barriers to entry as well as freedom of users to contribute freely and create communities around shared interests (Mayfield, 2008). These platforms have broken down the geographical barriers that once restricted communication and have led to increased information sharing.

With profound benefits, it is easy to overlook the harms of social media, which are of increasingly important consideration due to the continued proliferation of social media platforms. Beyond the alarming headlines such as “Have Smartphones Destroyed a Generation?”¹ and “Social Media Linked to Rise in Mental Health Disorders”², there is a growing body of research demonstrating the destructive impact of social media not only on individuals but on societal welfare³.

Research Objectives and Aims

The objective of this thesis is to bring attention to the harms of social media on individuals and society. Through a comprehensive review of the current literature on social media, this thesis identifies and explores how the underlying architecture of social media platforms causes harm. This thesis also assesses the regulatory frameworks created to address the challenges that social media companies pose.

The aim of the analysis is, therefore, threefold:

1. To identify and explore the motivation behind the algorithms used by the dominant social media platforms.
2. To analyse the consequences of technology companies' leveraging human psychology. A review of the research identifies societal and individual outcomes of this phenomenon.

¹ Jean M Twenge. “Have Smart Phones Destroyed a Generation?” The Atlantic. September 2017. <https://www.theatlantic.com/magazine/archive/2017/09/has-the-smartphone-destroyed-a-generation/534198/>

² Shamard Charles. Social Media Linked to Rise in Mental Health Disorders. NBC News. March 4, 2019. <https://www.nbcnews.com/health/mental-health/social-media-linked-rise-mental-health-disorders-teens-survey-finds-n982526>

³ Centre for Humane Technology. Ledger of Harms. June 2021. <https://ledger.humanetech.com/>

3. To assess the regulation put in place to address the harms caused by social media platforms and to suggest possible solutions to the challenges created by social media.

Study Significance

This thesis offers a comprehensive review of the harmful effects of social media. The thesis contributes to the literature by identifying and providing evidence for the range of externalities caused by the ubiquity of social media use. This thesis provides information that could help users and regulators maximise the benefits of social media platforms while mitigating the profound negative impact they cause in the long run. The assessments of frameworks used to regulate social media platforms could prove helpful in informing policy when establishing guardrails for virtual worlds, mainly social media platforms.

Chapter 2: Method

This thesis aims to comprehensively review the evidence regarding the effects of social media on individual and societal well-being, grounding its methodology in systematic reviewing techniques (McFadden et al., 2012) and include seeking transparent and rigorous approaches to identification, quality appraisal and synthesis of studies. The adopted approach prioritises transparent and rigorous methods for identification, quality assessment, and synthesis of studies, aiming to present a reliable depiction of the current best evidence relevant to the inquiry (MacDonald, 2003).

Systematic reviews, are crafted to offer a dependable portrayal of existing evidence pertinent to a specific query, placing substantial emphasis not only on selection and appraisal methods but also on the systematic quality of data synthesis (Killick & Taylor, 2009). As outlined by Campbell et al. (2003), 'synthesis' encapsulates the process of extracting, interpreting, and representing data from individual research studies in a collective manner.

Given the nature of the research question and the varied research designs involved in this review, the feasibility of conducting a statistical meta-analysis was limited. Therefore, a thematic analysis approach was employed for synthesis. Thematic analysis provides a method to explore, identify, and analyse patterns of meaning, or themes, within the literature, guided by the framework established by Braun and Clarke (2006).

Furthermore, recognising the dynamic nature of the internet as a rapidly evolving phenomenon, this study did not confine its exploration solely to academic sources. In discussing the various identified themes, non-academic sources, such as news articles and contemporary commentaries, were also incorporated. This inclusion aimed to capture the dynamic nuances and ongoing developments surrounding the identified themes while considering both academic research and real-time developments in the digital landscape.

Search Strategy

This study utilised systematic searching techniques to retrieve relevant research studies pertaining to the search topic (McFadden et al., 2012). A systematic literature search was conducted, seeking relevant articles. Social media were defined as websites used primarily for social interaction: these include social networking sites such as Facebook, instant messaging (e.g., WhatsApp) and image and video sharing applications such as Instagram. No time period

was applied, or geographical restriction; such limits would be arbitrary and a risk of selection bias (McCrae and Purssell 2015). Articles were screened first by title, then by abstract and at the final stage of screening, full articles were read and assessed.

Searches were performed on the following databases: Google Scholar, Web of Science, JSTOR, PsycINFO and Scopus. The flow diagram below gives an overview of the systematic search strategy.

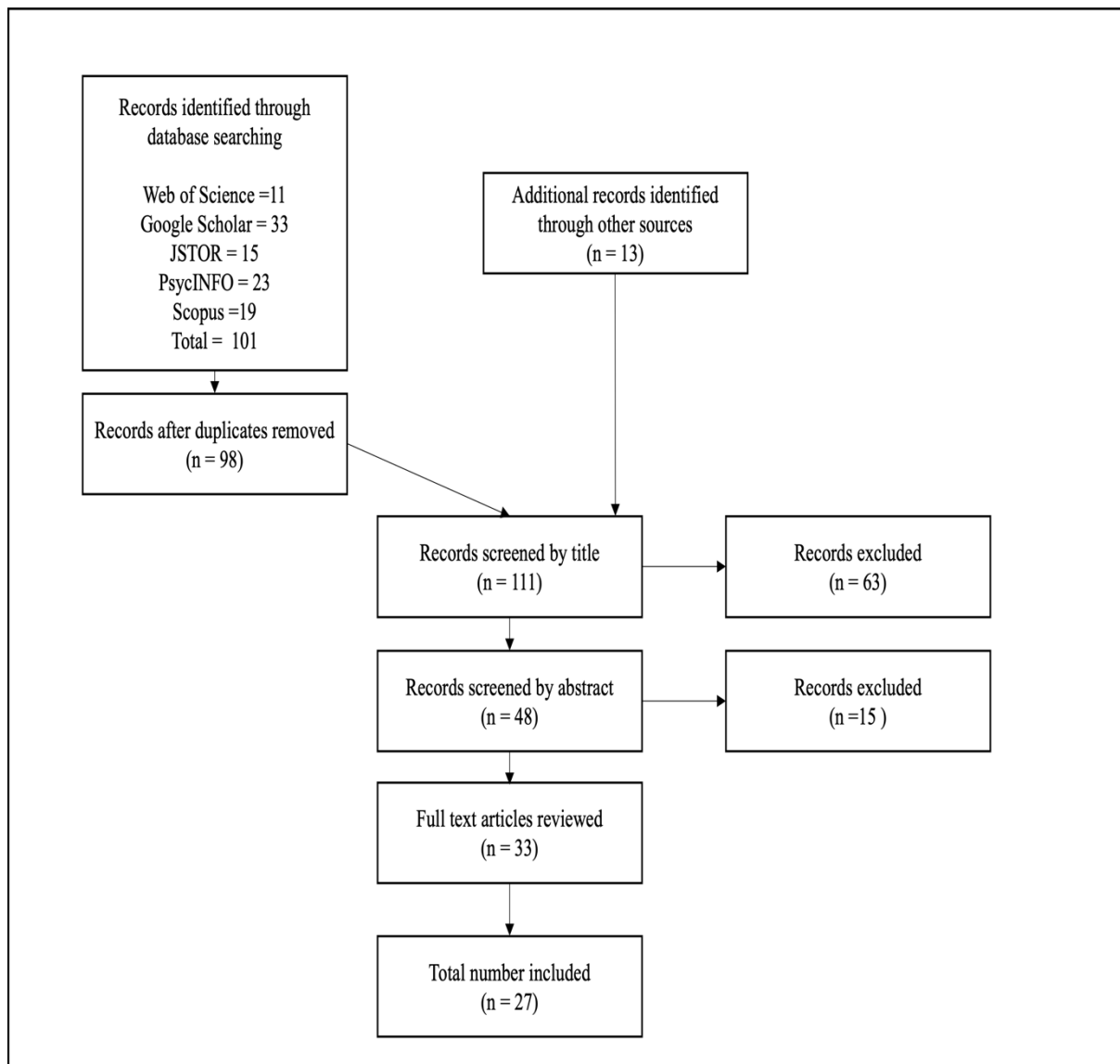


Figure 1. Overview of the systematic search strategy used in this thesis.

Selecting for relevance

A focused lens was applied to select studies that predominantly addressed the nuanced facets of social media's influence. The selection criteria for inclusion in this study centred on the direct correlation of the literature with various communicative social media technologies. This included blogs, message boards, interactive websites, forums, social networking sites and video sharing platforms, thus, ensuring a comprehensive examination of diverse platforms. Grey literature and non-English language papers were excluded due to limitations in time and cost. Additionally, papers focusing solely on the internet's general impact were omitted unless they explicitly delved into variables pertaining to interactive online communication with others. As social media platforms are also essential drivers of smartphone usage, studies involving smartphone usage were also included. A total of 101 studies were identified and subject to full text review. Subsequent removal of duplicate studies, theoretical material, descriptive case studies, and policy documents resulted in a refined selection of 27 studies that formed the backbone of this investigation. Due to the exploratory nature of this review, more relevant studies and articles were included while researching the individual themes.

Table 1 Summary of Studies

Authors	Country	Internet Issues	Methods	Key Findings
Vosoughi et al. (2018)	US	Misinformation	Quantitative Experiment	Falsehood spread significantly farther, faster, deeper, and more broadly than true stories across all categories of information.
Howard et al. (2018)	US	Misinformation	Quantitative	Fake news items tend to contain higher emotional content and present unexpected information, making them more likely to be shared and reposted extensively.
Bessi et al. (2016)	UK	Polarisation	Quantitative	Content drives the emergence of echo chambers on both YouTube and Facebook.
Matakos et al. (2017)	Greece US	Polarisation	Quantitative Experiment	Polarisation in society, especially amplified through the widespread use of online social networks
Boxell et al. (2017)	US	Polarisation	Quantitative	Polarisation has increased the most among the demographic groups least likely to use the Internet and social media, suggesting that the role of these factors is limited.
Di Domenico et al. (2021)	UK	Fake News	Systematic Review	Social media plays a role in spreading misinformation, widely described as 'Fake News'
Lazer et al. (2018)	US	Fake News	Qualitative	Addressing fake news requires a multidisciplinary effort to understand how the Internet spreads content and how people process news.
Aral and Eckles. (2019)	US	Election manipulation	Qualitative	Rigorous causal analysis could help identify social media manipulation of elections
Müller and Schwarz (2021)	Germany	Political and Social Violence	Quantitative	Anti-refugee sentiment on Facebook predicts crimes against refugees in otherwise similar municipalities with higher social media usage.
Bradshaw and Howard (2019)	UK	Political manipulation	Qualitative	Social media facilitates the spread of political propaganda, suppresses freedom of speech, and pollutes the digital information system.

Authors	Country	Internet Issues	Methods	Key Findings
Ward et al. (2017)	US	Cognition	Experiment	The mere presence of one's own smartphone reduces available cognitive capacity
Hadar et al. (2017)	Israel	Cognition	Experiment	Heavy usage was found to be associated with impaired attention, reduced numerical processing capacity, changes in social cognition, and reduced right prefrontal cortex (rPFC) excitability.
Hutton et al. (2020)	US	Cognition	Quantitative	Screen use greater than that recommended by the American Academy of Pediatrics guidelines was associated with (1) lower measures of microstructural organisation and myelination of brain white matter tracts that support language and emergent literacy skills and (2) corresponding cognitive assessments.
Madigan et al. (2019)	Canada	Cognition	Quantitative	Higher levels of screen time in children aged 24 and 36 months were associated with poor performance on a screening measure assessing children's achievement of development milestones at 36 and 60 months, respectively.
Kushlev and Dunn (2019)	US	Attention	Experiment	Smartphones can distract parents from reaping a sense of social connection when spending time with their children. These studies suggest that being constantly connected to the Internet may carry subtle costs for the fabric of social life.
Kross et al. (2021)	US	Well-being	Systematic Review	Accumulating evidence indicates that social media can enhance or diminish well-being depending on how people use them.
Rae and Lonborg (2015)	US	Well-being	Quantitative	Quantity and motivations of Facebook use interact with potentially serious implications for psychological well-being

Authors	Country	Internet Issues	Methods	Key Findings
Boers et al. (2019)	Canada	Well-being	Longitudinal Quantitative	Time-varying associations between social media, television, and depression were found, which appeared to be more explained by upward social comparison and reinforcing spirals hypotheses
Viner et al. (2019)	UK	Well-being	Systematic review	There is evidence that higher levels of screentime is associated with a variety of health harms for children, with evidence strongest for adiposity, unhealthy diet, depressive symptoms and quality of life.
Coyne et al. (2020)	US	Depression Anxiety	Longitudinal Quantitative	Increased time spent on social media was not associated with increased mental health issues across development when examined at the individual level.
Twenge and Campbell (2019)	US	Well-being	Longitudinal Quantitative	There is a significant correlation between heavy usage of digital media (5+ hours a day) and lower psychological well-being.
Mosquera et al. (2020)	US	Depression	Experiment	Taking a break from Facebook for a week can have several positive effects. Firstly, individuals who abstain from Facebook tend to consume less news overall and are less likely to identify politically biased news stories.
Allcott et al. (2020)	US	Life Satisfaction	Experiment	Temporary breaks from Facebook may lead to improved mental well-being, a change in leisure activities, a more balanced news perspective, and a sustained decrease in overall Facebook usage.
Tromholt (2016)	Denmark	Mood	Experiment	It was demonstrated that taking a break from Facebook has positive effects on the two dimensions of well-being: our life satisfaction increases, and our emotions become more positive.
Giumetti and Kowalski (2022)	US	Cyberbullying	Systematic review	Negative effects on well-being that may be linked with cyberbullying via social media, including psychological distress, decreased life satisfaction, and suicidal ideation.

Developing a theoretical model: Thematic Analysis

Developing a theoretical framework required reviewing and categorising the studies to extract essential elements for the synthesis process. Derived from successful methodologies used in comparable studies (Killick & Taylor, 2009), this approach proved crucial, especially given the diverse range of knowledge fields contributing to the literature. Employing thematic analysis for each study allowed for a deductive approach in identifying and organising key themes and critical issues. As no single theory or model adequately covered the extensive scope of the literature, the synthesis necessitated a multi-dimensional analytical framework. This framework pulled together ideas from communication, sociology, and psychology, giving us a comprehensive way to understand how social media affects both individuals and society.

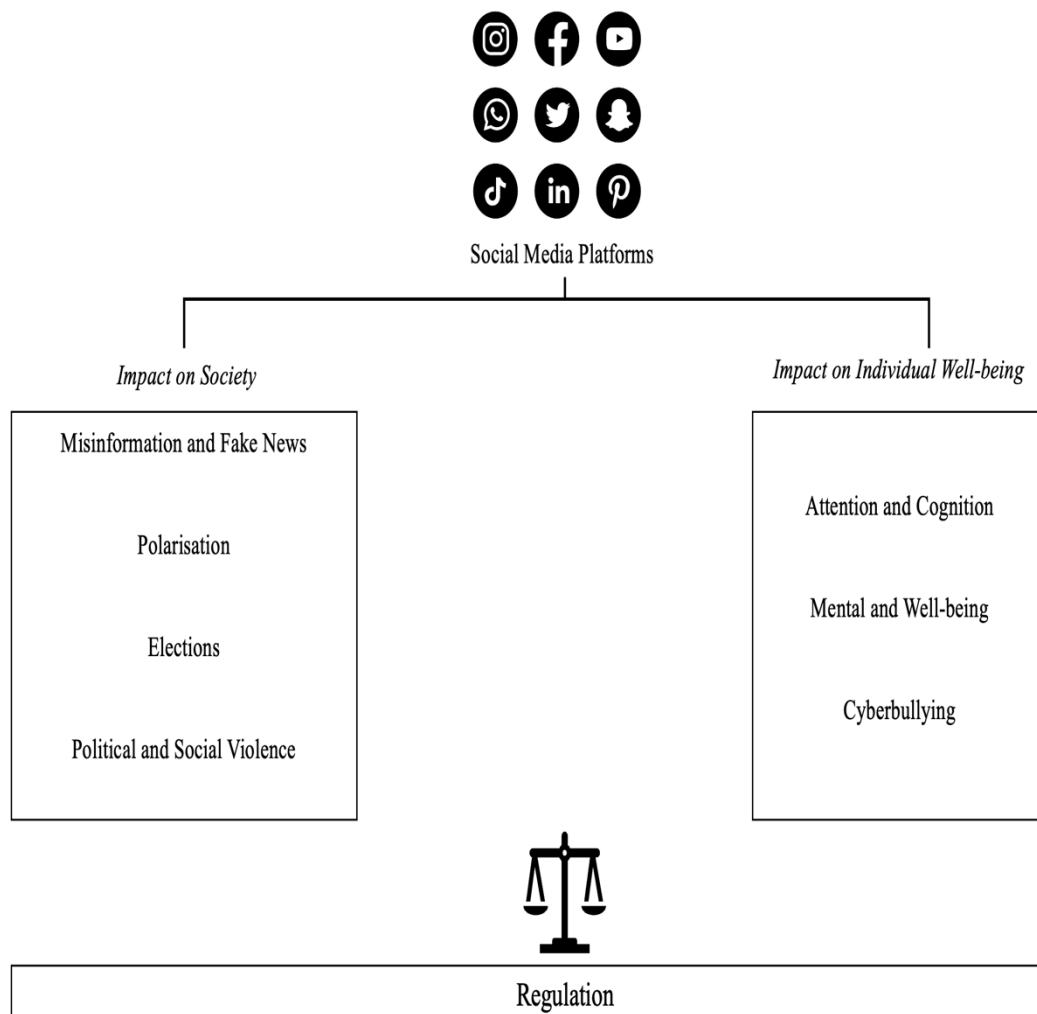


Figure 2. Theoretical themes on the impact of social media

In Figure 2, the identified theoretical strands of literature in this thesis are illustrated. The social media icons symbolize the diverse array of social media platforms available. The 'Implications on society' strand encompasses literature examining how social media affects society, while 'Impact of social media on individuals' focuses on research exploring its influence on personal well-being. It's noteworthy that researchers often use the term 'well-being' to denote various constructs or variables. Additionally, the 'regulation' theme initiates discussions on actionable measures.

Chapter 3: The Rise of Social Media

The objective of this chapter is to interrogate the motivations behind the algorithms social media companies use in their applications. The chapter begins by detailing the positive aspects of social media platforms that contribute to their popularity. The chapter then discusses how social media companies use the advertising model to make money off their platforms. Lastly, the chapter interrogates the design of social media platforms, illustrating how technology companies leverage human psychology for corporate greed using persuasive design techniques.

At the core of the evolution of social media is the human desire for connection. The early years of Facebook, the most popular social media site, were filled with wholesome interactions and positive stories, such as that of long-lost siblings finding each other through the platform⁴. Facebook's mission to "bring the world closer together"⁵ exemplifies this desire for interconnectedness. With more than two billion users worldwide, the company is seemingly on track to achieving its mission. Platforms like Facebook and Twitter act as virtual gathering places, bringing together individuals and ideas from around the world. They provide a space where users with shared interests can find or create community enabling collaboration at revolutionary scales.

The influence of social media extends beyond cyberspace. Social media has been at the centre of many social justice movements, used to raise awareness for different causes and as a communication tool, spreading important coordination information. For instance, social media played an essential role in organising the Arab Spring, in which pro-democracy protestors in North Africa and the Middle East rebelled against sitting governments. Studies show that social media enhanced the scope and efficacy of the movement and provided sources of information beyond governments' control (Tufekci and Wilson, 2012). Globally, activists utilise social media for mobilisation and organisation, from protest movements like Black Lives Matter, which highlights racism experienced by Black people, to the #Me Too movement, created to speak out against the sexual harassment of women. One cannot overstate social media's crucial role in generating awareness and disseminating information.

⁴ Gillian Mohny. "Long-Lost siblings Find Each Other on Facebook" ABC News, October 28, 2014. <https://abcnews.go.com/US/long-lost-siblings-find-facebook/story?id=26514449>

⁵ Meta. Facebook's Mission. <https://about.facebook.com/company-info/>

In the face of the COVID-19 pandemic, social media again became vital in disseminating regional, national, and international information. Never in human history has it been possible to communicate as easily and rapidly during a pandemic. Social media enabled the sharing of new important information, relevant scientific findings and a means to compare different approaches globally (González-Padilla and Tortolero-Blanco, 2020). For instance, an infographic about airway management of patients with COVID-19 was shared on Twitter. The post went viral, leading to the translation of the infographic into more than ten languages due to multiple requests and the relevance of the information (González-Padilla and Tortolero-Blanco, 2020). This educational piece of information was shared across geographic boundaries and adapted according to each healthcare setting. Furthermore, during this time, social media platforms became essential for maintaining communication between friends and family, reducing feelings of isolation and loneliness and giving people a sense of community

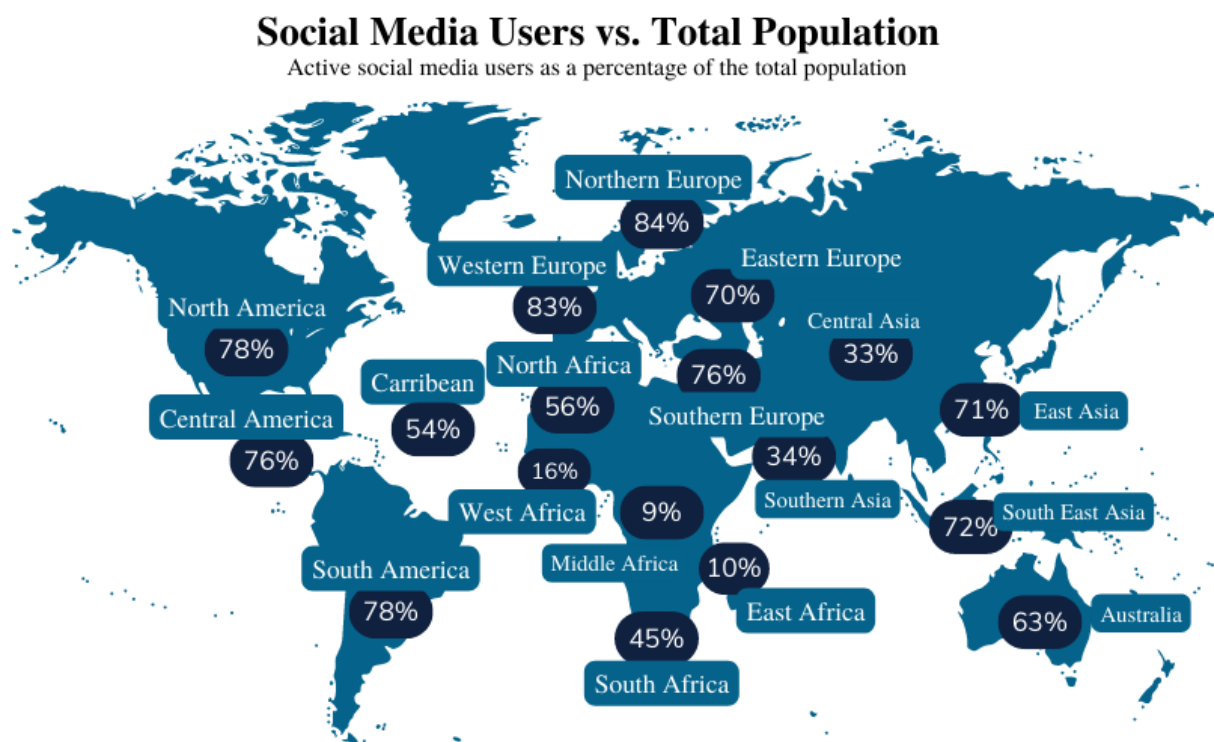


Figure 3. Social media global penetration. The figure illustrates the number of active social media users as a percentage of the total population of the country. (Source: Hootsuite, July 2022.)

However, the very qualities that make social media invaluable also contribute to the darker sides of these platforms. Social media's hyperconnectivity contributes to its addictive nature, negatively impacting users' psychological well-being. The anonymity and unity that the platforms provide are essential in social justice movements, but they also enable hate speech

and racialised violence towards marginalised groups. The ability to rapidly disseminate educational information during the COVID-19 pandemic also promoted the proliferation of misinformation to such an extent that the World Health Organisation stated that the misinformation on social media was as threatening to global public health as the virus itself⁶.

The negative impacts of social media are so profound that founders and employees at social media companies have begun speaking out. For instance, Chamath Palihapitiya, former vice president of user growth at Facebook, stated:

“The short-term, dopamine-driven feedback loops we have created are destroying how society works.... No civil discourse, no cooperation, misinformation, mistruth. And it’s not an American problem—this is not about Russian ads. This is a global problem. I feel tremendous guilt. I think we all knew in the back of our minds—even though we feigned this whole line of, like, there probably aren’t any bad unintended consequences. I think in the back, deep, deep recesses, we knew something bad could happen. So, we are in a really bad state of affairs right now, in my opinion. It is eroding the core foundation of how people behave by and between each other. And I don’t have a good solution. My solution is I don’t use these tools anymore. I haven’t for years”⁷.

Furthermore, Sean Parker, the first President of Facebook, states:

“We need to sort of give you a little dopamine hit every once in a while, because someone liked or commented on a photo or a post or whatever. It’s a social-validation feedback loop, exactly the kind of thing that a hacker like myself would come up with, because you’re exploiting a vulnerability in human psychology. The inventors, creators—it’s me, Mark [Zuckerberg], Kevin Systrom on Instagram, all of these people—understood this consciously. And we did it anyway; it changes your relationship with society, with each other. It probably

⁶ World Health Organization. “Social media & COVID-19: A global study of digital crisis interaction among Gen Z and Millennials”. December 1, 2021. <https://www.who.int/news-room/feature-stories/detail/social-media-covid-19-a-global-study-of-digital-crisis-interaction-among-gen-z-and-millennials>

⁷ James Vincent. “Former Facebook exec says social media is ripping apart society.” The Verge. December 11, 2017. <https://www.theverge.com/2017/12/11/16761016/former-facebook-exec-ripping-apart-society>

interferes with productivity in weird ways. God only, knows what it's doing to our children's brains.”⁸

It is undeniable that social media is causing harm, primarily due to the underlying architecture of the platforms that are ultimately designed for corporate profit.

The Attention Economy and the Advertising Revenue model

We live in the Age of Information, in which social media, television, the internet and news sources constantly bombard people with a never-ending stream of content. Amidst the mass of information available, it is difficult for one source to captivate attention and be influential. Cognitive psychologist Herbert Simon observed this in 1971 when he wrote, “a wealth of information creates a poverty of attention.” Today, human attention is a scarce resource and harvesting it has become an essential part of the revenue generation strategy for many businesses, including social media platforms⁹.

Social media platforms capitalise on commodifying human attention using the advertising revenue model (Mahoney and Tang, 2016). Social media companies seek to captivate and maximise user attention to sell to their customers — advertisers. Advertisers may include large corporations like Coca-Cola or Samsung. The advertisers bid for the ability to direct their ads toward users on social media platforms, paying billions of dollars in exchange. The more views or “eyeballs” a page collects, the more valuable it is for the advertising company and the more lucrative it is for the social media platform (Flores, 2022).

The advertising revenue model is the most common business model among social media networks (Mahoney and Tang, 2016). For instance, Meta, Facebook's parent company, generated \$117.9 billion in 2021. Advertisements generated 97.79% of that revenue (Statista, 2022). Having advertisements as the primary revenue source incentivises social media platforms to maximise user engagement at any cost. Content with a high degree of engagement has many people viewing it, creating a large audience for advertisers and an income stream for

⁸ Mike Allen. “Sean Parker unloads on Facebook: “God only knows what it's doing to our children's brains.” Axios, November 9, 2017. <https://www.axios.com/2017/12/15/sean-parker-unloads-on-facebook-god-only-knows-what-its-doing-to-our-childrens-brains-1513306792>

⁹ Centre for Humane Technology. “The Attention Economy.” <https://www.humanetech.com/youth/the-attention-economy>

social media companies (Herrick, 2022). As such, social media companies design their algorithms to maximise engagement and corporate profit.

Traditional advertising on TV, in newspapers and in magazines differs from social media advertising because it does not provide precise data to advertisers about the people looking at them. Now, using artificial intelligence, companies can measure individuals changed behaviours and have the platforms’ feed altered to influence user behaviour through user-related advertising. This type of advertising exploits user data to personalise the adverts shown and increase the conversion rates (Bühler et al., 2015). User-related advertising utilises data such as the user’s location, browser history and social media activities. The user’s social media activities are tracked, recorded, and registered into a social graph which provides advertisers with information.

Social media, therefore, utilises precise data to accurately predict and personalise content such that these platforms can influence user behaviour (Lanier, 2018). They collect in-depth data about how to influence user decisions and then sell that influence to the highest bidder. Advertisers can then target specific messages to a particular audience. While social media platforms are freely available to users, the users’ attention becomes the product sold by the platforms to advertisers. The more engaged the users are with the platform, the higher the profits social media companies make.

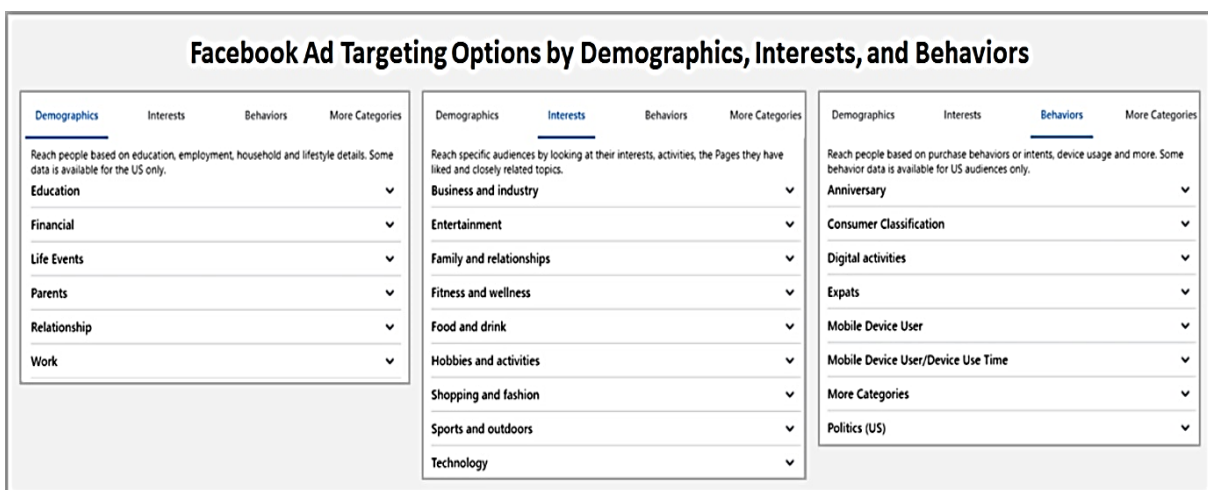


Figure 4. Examples of the primary targeting options Facebook offers to advertisers by demographics, interest, and behaviours. Source: Centre for Humane Technology, 2022

Social media companies prioritise engagement and growth by using persuasive design techniques. The maximisation of attention is not only harmful to the users but has led to considerable detrimental societal outcomes¹⁰. The advertising revenue model incentivises tech platforms to optimise engagement, despite the costs to society.

Persuasive Technology

Social media companies utilise designs that exploit the human reward system to keep individuals engaged in their services (Flores, 2022). The reward system consists of vital brain structures that determine behaviour and whether individuals form a habit. The reward system in our brains drives us toward stimuli such as food and away from painful stimuli such as conflict. (Lanuza et al., 2008). Platforms such as YouTube, Facebook, Twitter, and Instagram use persuasive technology explicitly designed to change users' opinions, attitudes, or behaviours to meet the companies' goals¹¹. Social media companies design features on their applications to persuade users to spend more time on their applications. Features such as push notifications, 'like' buttons and feed personalisation are all designed to incentivise users to stay on the platforms longer.

Artificial Intelligence is used in social media applications to personalise the content that users see. Furthermore, Persuasive Technology companies specialise in algorithms that influence human behaviour and sell that influence to advertisers. For instance, ByteDance is an Artificial Intelligence company that owns TikTok and several other applications worldwide. Their applications have sophisticated algorithms that make their business successful. Their algorithms study how people use TikTok, considering everything about their users; the websites they browse, their typing, and their keystroke rhythms and patterns (Centre for Humane Technology, 2021). These algorithms have made ByteDance one of the most valuable start-ups in the world, with 1.9 billion monthly active users across all its platforms as of June 2021 (Lin, 2017).

Advanced algorithms monitor and compare users' behaviour with the behaviour of others to discover how to influence them best. Applications then sell that access to companies or

¹⁰ Centre for Humane Technology. "The Attention Economy." <https://www.humanetech.com/youth/the-attention-economy>

¹¹ Centre for Humane Technology. "Persuasive Technology." <https://www.humanetech.com/youth/persuasive-technology>

individuals who wish to influence behaviour, opinions, or votes. Persuasive technology constantly learns more about users and pairs that information with compelling and creative design ideas to influence behaviour more effectively (Centre for Humane Technology, 2021). Social media companies have reported on their ability to influence their users' moods (Kramer et al., 2014), change voter turnout (Corbyn, 2012) and reinforce brand loyalty all by making changes to their algorithms (Lanier, 2018).

Conclusion

Social media companies design their algorithms to amplify popular content. Popular content captures attention that social media companies capitalise on through advertisements. Furthermore, technology companies use persuasive techniques to keep users on the platforms longer to capture their attention. Valuing user engagement regardless of the content amplifies negativity such as misinformation, sensationalism and hate speech. While some of these companies have claimed that they are only objective platforms and not information publishers¹², the reality is that the platform constantly shapes, influences, and defines the experience of everyone engaging with it through the algorithms they utilise. Social media companies' business model hinges on selling human attention and behaviour. This approach has resulted in several unintended consequences, including increased feelings of loneliness and anxiety among individuals who excessively engage online, as well as the proliferation of misinformation that undermines trust in institutions and fosters division within communities on crucial issues like elections—themes explored within this thesis.

¹² Jeffrey Herbst. "The Algorithm is an Editor; Google, Facebook and Other Tech Companies Say They Aren't News Organizations, but the Claim is Becoming Increasingly Implausible,". Wall Street Journal. April 13, 2016. <https://www.proquest.com/docview/1780643099?accountid=12205&forcedol=true>

Chapter 4: Societal Implications of Social Media

The objective of this chapter is to interrogate how social media companies alter the political and social landscape, distorting the collective psyche and facilitating the amplification of systematic oppression. Social media companies' current business model prioritises maximising user engagement at all costs. This chapter details those costs beginning with social media's role in disseminating misinformation. Social media characteristics, including the low barriers of entry and the rampant polarisation on the platforms, facilitate the spread of fake news leading to real-world consequences such as mob lynchings in India. The chapter then dissects the use of social media to manipulate the outcomes of elections and highlights the role social media plays in amplifying political instability, illustrated by the genocide in Myanmar.

Misinformation

Social media platforms are considered reliable sources of information. In many countries worldwide, most people receive news from social media (Shearer and Gottfried, 2017). A Reuters Institute study suggests that 51% of people with access to the internet use social media as a primary news source¹³, such that social media platforms provide information for even professional news articles. For instance, research shows that journalists deem tweets as newsworthy as headlines from the Associated Press, a global news organisation (McGregor and Molyneux, 2020). However, social media platforms harm the global information ecosystem by facilitating the rapid spread of misinformation and fake news.

Fake news refers to information that makes a claim about the world that is factually incorrect but not obviously so, such that there is some uncertainty about its accuracy (Aymanns et al., 2017). Fake news overlaps with other information disorders such as misinformation, disinformation and mal-information. Misinformation is information which is false or misleading. Disinformation is false information purposely spread to deceive and mal-information is information, which is based on reality, but used to inflict harm on a person, organisation or country for instance, a report that reveals a person's sexual orientation without justification (Lazer et al., 2018). Fake news is created and distributed to generate revenue or

¹³ The Reuters Institute for the Study of Journalism. "Digital News Report." <https://www.digitalnewsreport.org/>

manipulate the public’s perception of reality (Apuke and Omar, 2021). Although fake news is not new, the rapid rate at which it spreads is now of global concern (McGonagle, 2017).

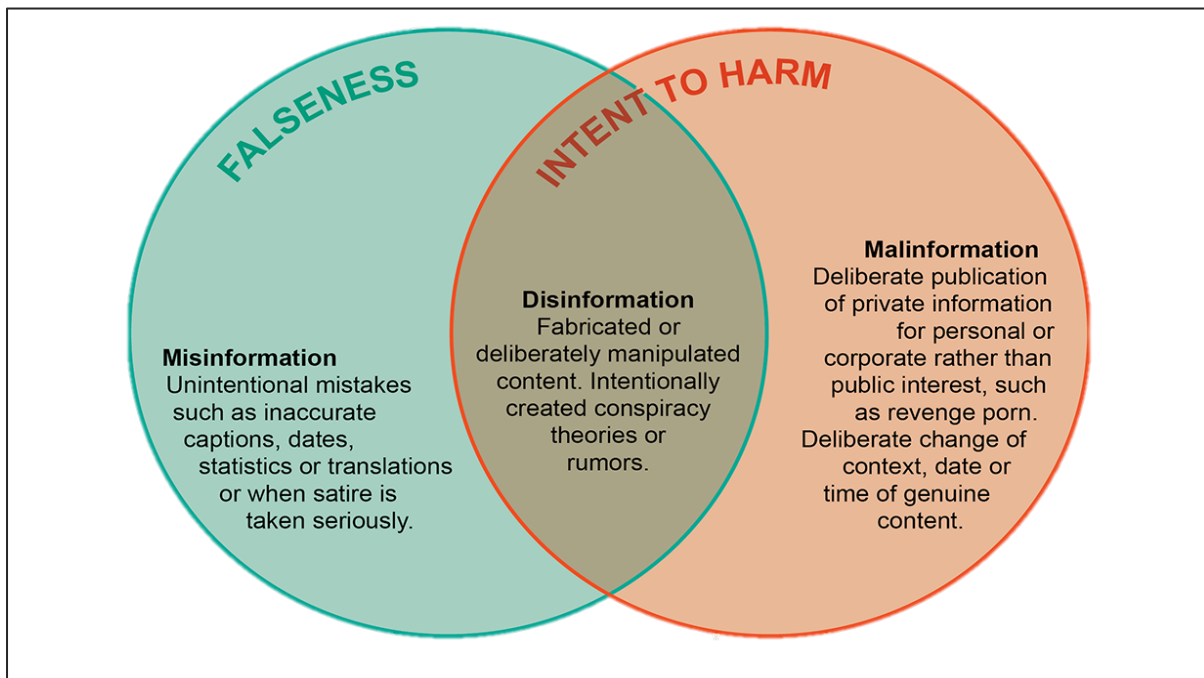


Figure 5. Categories of the Information Ecosystem. (Source: Scientific American, 2019)

Digital technologies like social media facilitate the rapid dissemination of information without filtering. Social media users can advance any ideas and spread news widely by simply liking, sharing, commenting, and retweeting. Lazer et al. (2018) found that social media platforms are influential for distributing a large amount of unfiltered content; as such, these platforms play a significant role in the proliferation of misinformation.

A study by Vosoughi et al. (2018) found that false information on Twitter spreads six times more rapidly than true stories, and falsehoods are 70% more likely to be retweeted than the truth. The researchers state that this may be because fake news is more attention-grabbing than authentic information. Fake news items usually have higher emotional content and contain unexpected information, which inevitably means that they will be shared and reposted more often. Similarly, a study of 22 million tweets showed that Twitter users shared more information characterised by ideological extremism, misinformation, and conspiratorial content than actual news stories (Howard et al., 2018).

Social media provides an ideal and conducive environment for spreading fake news (Lazer et al., 2018, Di Domenico et al., 2021, Zhang and Ghorbani, 2020) because of three main characteristics. The first characteristic is that there are *low barriers to entry* on social media.

The cost of entering the market and producing content is small (Allcott and Gentzkow, 2017). Furthermore, accounts can buy followers (both real and bots) and publish content to build their reputations at relatively low costs (Di Domenico et al., 2021).

The second feature is the *formatting* of information on social media. Information on social media is formatted in small slices to grab users' attention (Allcott and Gentzkow, 2017), so it is difficult to judge the accuracy of the information. The headline is deemed more important than the source of information, which means that what the user thinks about the news is more influential on the believability of the information than its source due to confirmation bias (Kim and Dennis, 2019).

The third feature is *polarisation*. Users on social media platforms read, interact, and share information that is consistent with their beliefs (Bessi et al., 2016, Di Domenico et al., 2021). Because of this, users create echo chambers. Algorithms on social media facilitate the creation and growth of echo chambers, and disinformation rapidly spreads through them because the information in these communities gets accepted even though it contains deliberately false information (Bessi et al., 2016, Di Domenico et al., 2021).

Social media platforms are also vulnerable to bots that facilitate the dissemination of misleading information. Bots are computer algorithms designed to impersonate human behaviour, interact with humans, and create platform content (Ferrara et al., 2016). Bots accelerate the spread of fake news and misinformation by artificially commenting, sharing or liking content (Lazer et al., 2018). According to Ferrara et al. (2016), bots are increasingly becoming 'smarter' and can create individual accounts and post material scouted from the internet at a scheduled rate. Bots play a significant role in spreading fake news on social media by amplifying the fake news in the initial stages to ensure the phoney narrative goes viral (Azzimonti and Fernandes, 2018). They also target influential users through mentions so that they may help facilitate the spreading process through their large follower base (Shao et al., 2018). Research states that approximately 9% to 15% of accounts on Twitter are social bots, while Facebook estimated that 60 million accounts on the platform are bots (Lazer et al., 2018).

Consequences of Fake News

This rapid spread of false information through social media is dangerous because it manipulates the public's perception of reality. Research shows that fake news items read even once increase

the probability of a reader judging that it is true when they next encounter it (Pennycook et al., 2018). Furthermore, explicitly warning individuals that third-party fact-checkers had disputed the fake news does not deter or even significantly diminish the reader from thinking the information is accurate, even if the news is counter to the reader's political standpoint (Pennycook et al., 2018).

Misleading information is harmful because it leads people to believe in information that is not true and subsequently base their decision on that false information (Lewandowsky et al., 2012). This creates a global challenge and undermines society's information ecology. For instance, false information regarding health issues threatens public health (Apuke and Omar, 2021). Research shows that spreading fake news and misinformation on social media is a primary cause of vaccine hesitancy and significantly affects immunisation choices (Carrieri et al., 2019, Aquino et al., 2017). In many countries, the consequences of fake news are even more fatal and lead to violence and death. For instance, in India, fake news forwarded through the Facebook-owned service WhatsApp led to the death of at least seventeen people through mob lynchings¹⁴.

Politically, fake news can influence voters. False political news travels more broadly and reaches more people than any other category of false information (Vosoughi et al., 2018). The extent to which fake news affects electoral outcomes is debatable (Van Duyn and Collier, 2019); despite this, fake news does affect the evaluation of news media and drives individuals to consume less political information (Mitchell et al., 2019). Voters' exposure to media topics on fake news negatively affects their media trust and impacts their ability to distinguish real from fake news about political candidates (Van Duyn and Collier, 2019). Fake News also leads to more extremism and polarisation among voters (Lazer et al., 2018).

The business model of social media platforms relies on monetising attention. Thus, their algorithms promote engaging and trending content rather than quality verifiable content. Personalisation algorithms also facilitate the creation of echo chambers, and the applications are vulnerable to bots, allowing fake news to thrive on the platforms. Social media contribute

¹⁴ Elizabeth Dwoskin and Annie Gowen. "On WhatsApp, fake news is fast — and can be fatal." The Washington Post. July 23, 2018. https://www.washingtonpost.com/business/economy/on-whatsapp-fake-news-is-fast--and-can-be-fatal/2018/07/23/a2dd7112-8ebf-11e8-bcd5-9d911c784c38_story.html

to information disorders on a global scale. Furthermore, these echo chambers divide social media users into distinct groups, which causes rampant polarisation on social media platforms.

Polarisation

In recent years, Social media emerged as a primary platform for public discourse and an influence on information consumption (Bessi et al., 2016). Individuals use platforms like Facebook and Twitter to express opinions and discuss ideas. Researchers thought that this would lead to individuals who are more open to different ideas, cultures, and viewpoints, thereby making societies more democratic and diverse overall. Interestingly, social media has the opposite effect (Matakos et al., 2017, Baumann et al., 2020).

Users on social media tend to interact with, share and read information consistent with their prior beliefs (Di Domenico et al., 2021). Users also connect and create communities around a shared narrative creating echo chambers (Bessi et al., 2016). *Echo chambers* are environments that reinforce a belief, opinion or political standpoint about a topic because of repeated interactions with peers or sources with similar thoughts and attitudes (Cinelli et al., 2021). Another facilitator of echo chambers on social media is personalisation algorithms (Di Domenico et al., 2021). Algorithms utilised on social media platforms personalise content based on users' previous behaviours, tastes, and preferences. Consequently, the information users see on their timeline is the content they agree with or have already been exposed to, further reinforcing their beliefs and viewpoints and increasing polarisation (Bessi et al., 2016).

Polarised communities emerge around heterogeneous narratives about a topic on which there is often extreme disagreement (Bessi et al., 2016). Online polarisation occurs over various issues such as abortion, politics, climate change denial and vaccines (Tucker et al., 2018). Polarisation divides people into sides with little communication and understanding of each other (Baumann et al., 2020). This results in the radicalisation of attitudes, extremism, false perception of events and further spread of misinformation within these communities because dissenting voices are filtered out (Tucker et al., 2018).

Instagram Polarisation on UK Referendum Debate

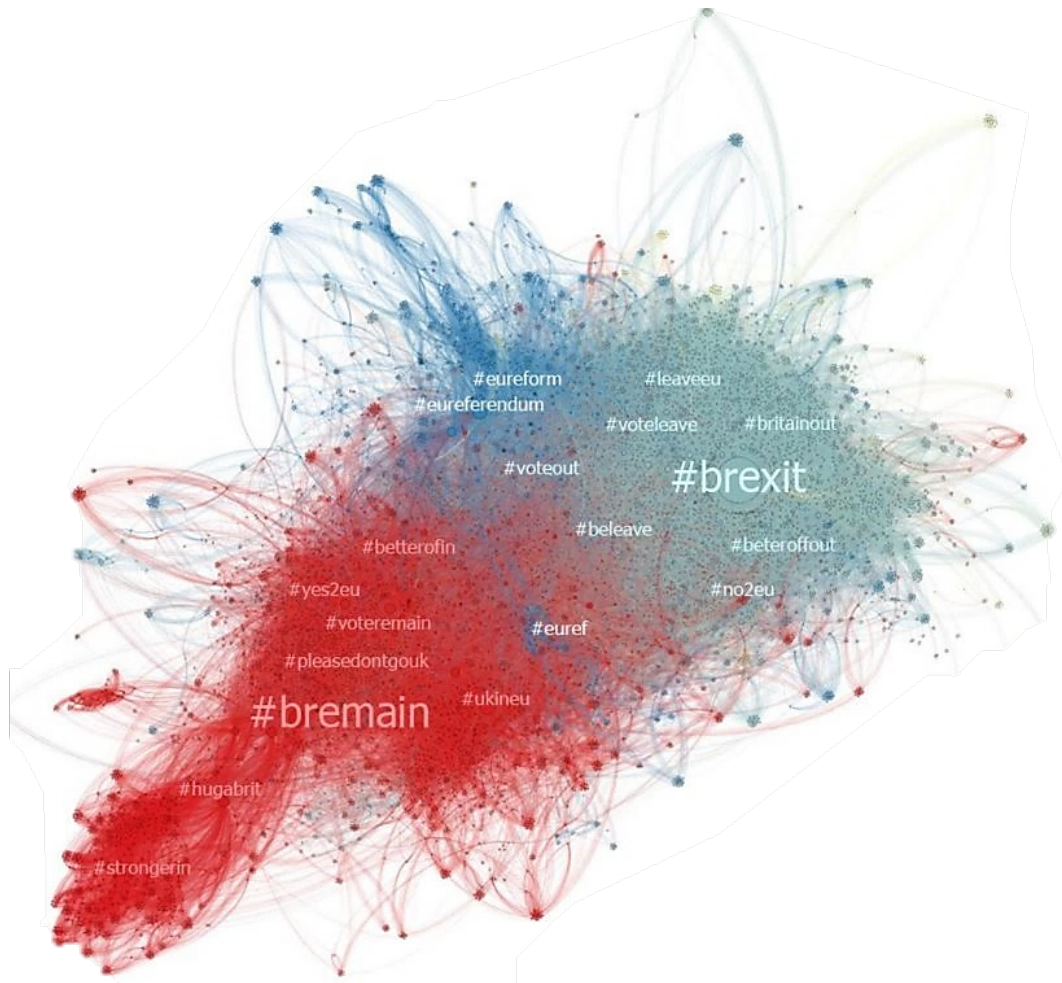


Figure 6. Network Analysis of hashtags on Instagram on the U.K's referendum to leave the E.U. (Source: Vyacheslav W. Polonski, 2016. Social Media Voices in the UK's EU Referendum)

Cinelli et al. (2021) found that polarised communities dominate online interactions on Facebook and Twitter and highlighted higher segregation on Facebook. Bessi et al. (2016) provide similar evidence of polarisation on YouTube. The study observed that most users who initially switch between content supporting different narratives eventually consume mainly one type of information, becoming polarised towards one of the two conflicting narratives. In contrast, Boxell et al. (2017) found that the increase in mass political polarisation has been most significant among citizens least likely to use the internet and social media. Their study shows that the internet explains only a small share of the recent growth in polarisation.

Despite the lack of consensus on the effect of social media on polarisation (Tucker et al., 2018), research shows that social media limits the exposure to diverse perspectives leading to extremism (Bessi et al., 2016), hate speech and real-life outcomes such as harassment and political violence (Lewis, 2018). The algorithms used by social media companies prioritise

viral engagement and thus amplify the most divisive content. YouTube is facing increasing scrutiny from scholars, media, and regulators because the video-sharing site seemingly guides users towards more extreme content no matter what users may be searching for¹⁵. A search on vegetarianism might lead to veganism; A search on Donald Trump's rallies might lead to white supremacist rants and holocaust denial videos. A quantitative study by Ribeiro et al. (2020) provides strong evidence of radicalisation pathways on YouTube. YouTube is extremely popular, especially among young people; a consequence of the website radicalising individuals could push harmful ideologies like white supremacy further into the mainstream.

The polarisation on social media often intensifies around election periods. Thus, exploring social media's role in elections is key.

Elections

A vital characteristic of a free and fair society is the ability to conduct a peaceful transition of power from one leader to the next. Social media plays an influential role in shaping events around elections. Unlike traditional news platforms like newspapers or radio, social media facilitates political discourse and enables voters and candidates to interact directly. Social media sites' increased reach and capacity have led to many questioning whether democratic elections are vulnerable to social media manipulation.

Over the last few years, there has been growing concern about the undue influence that social media platforms have on the outcome of elections and their impact on democracies around the globe (Tambini, 2018, Isaak and Hanna, 2018). The 2016 vote for the U.K. to leave the European Union and Donald Trump's victory in the 2016 U.S. Presidential election heightened this concern.

There is much disagreement amongst experts and researchers on whether social media manipulates voting decisions. Currently, minimal empirical evidence exists to substantiate the topic, mainly because of the distinctive challenges of studying social media's influence on election results (Aral and Eckles, 2019). Despite this, many argue that social media influences

¹⁵Zeynep Tufekci. "YouTube, the Great Radicalizer." The New York Times, March 10, 2018. <https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html>

the political environment and elections. One of the ways it does this is through the undermining of electoral legitimacy through persuasive techniques.

In the 2016 US presidential election, Hillary Clinton, Donald Trump, and Barack Obama¹⁶ argued that social media was instrumental in the election's outcome. As Brad Parscale, Trump's digital media director put it: "Facebook and Twitter were the reason we won this thing"¹⁷. Social media enabled Trump to direct targeted advertisements at voters in selected states. Donald Trump's digital team carefully tailored and tested Facebook advertisements to influence small segments of potential Trump voters to show up at the polls. In other cases, the social media advertisements targeted Clinton supporters, specifically white liberals, young women, and African Americans, with messages aimed at reducing turnout among those groups¹⁸. Social media companies such as Facebook use surveillance technologies to modify user behaviour for their customers, in this case, election candidates. A growing body of research has shown that social media can influence people's behaviour (Aral and Walker, 2012, Bond et al., 2012, Kramer et al., 2014). Facebook published research that showed it could change voter turnout (Corbyn, 2012). While the published study focused on boosting voter turnout, the platform could just as easily suppress targeted voters or potentially influence whom they vote for (Lanier, 2018).

In an attempt to understand Brexit- a referendum that went against voters' economic interests and Trump's victory in the US elections, a prospective explanation emerged; Cambridge Analytica. Cambridge Analytica was a political consulting company involved in the Trump campaign as well as the campaign to convince U.K. voters to choose to leave the European Union (Vaidhyathan, 2018).

¹⁶ David Remnick. "Obama reckons with a Trump presidency." The New Yorker, November 18, 2016 <https://www.newyorker.com/magazine/2016/11/28/obama-reckons-with-a-trump-presidency>

¹⁷ Lois Beckett. "Trump digital director says Facebook helped win the White House" The Guardian. The Guardian, October 9, 2017. <https://www.theguardian.com/technology/2017/oct/08/trump-digital-director-brad-parscale-facebook-advertising>

¹⁸ Joshua Green and Sasha Issenberg. "Inside the Trump Bunker, With Days to Go." Bloomberg, October 27, 2016. <https://www.bloomberg.com/news/articles/2016-10-27/inside-the-trump-bunker-with-12-days-to-go>

Cambridge Analytica built models to predict voter behaviour using personality trait data of voters and convinced campaigns in the United States and around the world that the models would help target and persuade voters¹⁹. Alexander Nix, Cambridge Analytica's CEO, stated that targeted advertisements using psychographic information would allow a firm or campaign to develop the right message for a particular voter or narrow set of voters²⁰. Nix claimed that by having hundreds of thousands of Americans' personality traits, Cambridge Analytica was able to form a model to predict the personality of every single adult in the U.S. In fact, after Trump's victory, in a press release, Nix stated, "We are thrilled that our revolutionary approach to data-driven communication has played such an integral part in President-elect Trump's extraordinary win."²¹ implying that Cambridge Analytica had the data to help Trump win.

Later, reports revealed that Cambridge Analytica mined the data of over eighty million Facebook users without their knowledge or permission²². Cambridge Analytica obtained the data from an external researcher who had claimed to Facebook that it was for academic purposes. Facebook banned Cambridge Analytica from advertising on its platform. Still, because the social media giant had no other measures to penalise companies that exploited data like that effectively, Facebook faced massive backlash.

Facebook received scrutiny from U.K and U.S politicians, regulators, and users, sparking the online movement #DeleteFacebook which trended on Twitter. Despite this, Facebook walked away relatively unscathed from the scandal. The social media company paid a \$5 billion penalty to the Federal Trade Commission to resolve the investigation into its privacy practices

¹⁹ Matthew Rosenberg, Nicholas Confessore, and Carole Cadwalladr. "How Trump Consultants Exploited the Facebook Data of Millions." The New York Times, March 17, 2018. <https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html>

²⁰ Alexander Nix. "The Power of Big Data and Psychographics." September 2016.

²¹ Cambridge Analytica Congratulates President-Elect Donald Trump and Vice President-Elect Mike Pence," November 9, 2016, <http://www.prnewswire.com/news-releases/cambridge-analytica-congratulates-president-elect-donald-trump-and-vice-president-elect-mike-pence-300359987.html>.

²² Matthew Rosenberg, Nicholas Confessore, and Carole Cadwalladr. "How Trump Consultants Exploited the Facebook Data of Millions." The New York Times, March 17, 2018. <https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html>.

and a £500,000 fine to the UK government, which is minor considering that the penalties equated to about a month of revenue for Facebook²³.

Scholars and Political scientists have been critical of the claims Cambridge Analytica made about the effectiveness of its methods²⁴. There is no clear evidence that Cambridge Analytica was successful in producing valuable models or that psychographic information performs better than other standard data-intensive techniques (Vaidhyanathan, 2018). Reports show that Cambridge Analytica exaggerated its capabilities, with executives admitting that the company never used psychographics in the Trump campaign²⁵. Additionally, the U.K. Information Commissioner's Office (ICO) found no evidence that Cambridge Analytica influenced the Brexit referendum in 2016²⁶.

The scandal of Cambridge Analytica made headlines and brought attention to many concerns around Facebook, data privacy and surveillance. However, it also obscured a dominant factor that did play a role in these elections- social media micro-targeting.

Social media companies control valuable personal data about voter preferences and behaviour. Research shows that Facebook can accurately calculate user political affiliation by tracking users' activity (Lieberman, 2016). With all the data social media companies control, a campaign would not need Cambridge Analytica if companies like Facebook could target voters themselves (Vaidhyanathan, 2018).

²³ Jaclyn Diaz. "Facebook's new whistle blower is renewing scrutiny of the social media giant" NPR. October 4, 2021. <https://www.npr.org/2021/10/04/1042921981/facebook-whistleblower-renewing-scrutiny-of-social-media-giant>

²⁴ Angela Chen and Alessandra Potenza. "Cambridge Analytica's Facebook data abuse shouldn't get credit for Trump". The Verge. 20 March 2018. <https://www.theverge.com/2018/3/20/17138854/cambridge-analytica-facebook-data-trump-campaign-psychographic-microtargeting>

²⁵ Nicholas Confessore and Danny Hakim. "Data Firm Says, 'Secret Sauce' Aided Trump; Many Scoff". The New York Times. March 2017. <https://www.nytimes.com/2017/03/06/us/politics/cambridge-analytica.html>

²⁶ Laurenz Gherkhe. "UK probe finds no evidence that Cambridge Analytica misused data to influence Brexit". POLITICO. 7 October 2020. <https://www.politico.eu/article/no-evidence-that-cambridge-analytica-misused-data-to-influence-brexit-report/>

In the 2016 US election, Facebook, Twitter, Microsoft, and Google were a part of all major presidential candidates' digital teams. Facebook and Google through YouTube were the most influential (Vaidhyanathan, 2018). Facebook had superior penetration across the American electorate, and because of its robust advertising targeting services, it became a powerful ally in the election. While the company had offered both the Clinton and Trump campaigns onsite consultants to help maximise their use of the platform, only Trump's campaign accepted the offer²⁷.

Political campaigns work with social media companies to harvest consumer data, profile voters based on topics of interest, and adjust their messaging for these narrow segments of potential voters (Vaidhyanathan, 2018). Using personal user data, social media companies can generate algorithmic tools that focus advertisements on voters most likely to be moved by specific messages (Lanier, 2018). The citizens profiled are unaware that companies are processing their information and targeting them. This lack of awareness takes away autonomy from citizens and brings up several problems.

Firstly, it contributes to redlining as segments of the population are encouraged to abstain from voting because they would likely vote unfavourably (Vaidhyanathan, 2018). Additionally, voters receive customised messages relevant to issues that concern them. Targeted messages motivate potential supporters and distract voters from other policies they might object to. Vaidhyanathan (2018) argues that campaigns like this result in citizens being "managed". This means that candidates can manipulate and precisely target information depending on the potential voters' needs of the moment. This overlooks the overall campaigns to conceal core policy positions or make them "strategically ambiguous" to get more votes.

Social media companies are not obligated to reveal what data they have or how they use it. Furthermore, the companies do not currently face regulatory restrictions on how data is used for persuasion, as social media companies are primarily in the business of selling advertisements. Thus, persuasion is their purpose.

Social media's influence in elections is not limited to the U.S and the U.K. Facebook assisted Prime Minister Narendra Modi in developing an online presence in India. In the Philippines, it

²⁷ Lesley Stahl. "Facebook "embeds," Russia and the Trump campaign's secret weapon." October 2017. <https://www.cbsnews.com/news/facebook-embeds-russia-and-the-trump-campaigns-secret-weapon/>

trained the campaign of Rodrigo Duterte, known for encouraging extrajudicial killings, on how to use the platform most effectively. In Germany, according to campaign staff, Facebook helped the anti-immigrant Alternative for Germany party (AfD) win its first Bundestag seats²⁸.

Facebook says it does not provide strategic advice but instruction on how best to use the platform. The company trains politicians and leaders on how to set up a campaign page, how to best use video to engage viewers and how to target ads to critical voting blocks. In exchange, once candidates are elected, their relationship with Facebook can help extend the company's reach into government in meaningful ways, such as being well positioned to push against regulations²⁸.

Beyond this, social media also affects elections through the rapid misinformation on its platforms. For instance, a study commissioned by the U.S. Senate Intelligence Committee reported Russia's use of social media to influence the 2016 US presidential election in favour of Donald Trump²⁹. The report identifies Russia's strategy to manipulate the polls using misinformation, paid social media advertisements and trolls. A year after the election, investigations revealed that one of the strategies Russian intelligence used was creating African American activist accounts. These accounts amplified political discord around issues like race and immigration and intended to suppress the vote statistically by lowering the enthusiasm of Black activists and Black voters³⁰. Social media company lawyers testified under oath that the companies were unaware that Russian intelligence services, through paid advertisements, sought to manipulate the US elections and foment divisions among voters.³¹

²⁸ Lauren Etter, Vernon Silver, and Sarah Frier. "How Facebook's Political Unit Enables the Dark Art of Digital Propaganda." December 2017. <https://www.bloomberg.com/news/features/2017-12-21/inside-the-facebook-team-helping-regimes-that-reach-out-and-crack-down>

²⁹ Office of the Director of National Intelligence. Background to "Assessing Russian Activities and Intentions in Recent US Elections: The Analytic Process and Cyber Incident Attribution." 2017

³⁰ Donie O'Sullivan and Dylan Byers. "Fake Black activist accounts linked to the Russian government." CNN, September 28, 2017. <https://money.cnn.com/2017/09/28/media/blacktivist-russia-facebook-twitter/index.html>

³¹ Hamza Shaban, Craig Timberg, and Elizabeth Dwoskin. Facebook, Google, and Twitter testified on Capitol Hill. Here's what they said. The Washington Post, October 31, 2017. <https://www.washingtonpost.com/news/the-switch/wp/2017/10/31/facebook-google-and-twitter-are-set-to-testify-on-capitol-hill-heres-what-to-expect/>

The U.S Senate Intelligence report highlights but does not answer whether social media manipulation may have influenced the outcome. Some researchers argue that the Russian-sponsored content was too minimal to influence the outcome (Allcott and Gentzkow, 2017), while others claim that Russia's influence likely tipped the election for Donald Trump (Jamieson, 2020).

Despite the disagreement amongst researchers, studies have shown that people tend to vote based on their recollection of candidates and campaign events, but these memories are not always accurate (Lau and Redlawsk, 2006). Fake news and misinformation from social media can influence voters' memories. A study by Murphy et al. (2019) found that exposure to a fake political news story can rewire voters' memories: in the study, researchers showed over 3,000 voters fake stories, and many voters later not only "remembered" the fake stories as if they were actual events but also "remembered" additional, rich details of how and when the events took place. The spreading of misinformation through social media can influence individuals and whom they vote for based on what they see online.

Regulatory bodies and governing institutions are grappling with regulating influential social media platforms in a meaningful way. This is heightened by the fact that regulators cannot anticipate social media's future impact on the rights and duties of citizens. This is alarming because unethical technological practices by social media companies can have disastrous and far-reaching consequences on democratic processes like elections worldwide.

Political and Social Violence

The effects of social media not only affect its users but spill over into the broader society, affecting non-users. Fake news, misinformation, polarisation, and election manipulation facilitated through social media affect countries' political and social landscapes. While many activists have used social media to organise protests, spread awareness and defend democracies for over a decade, social media has also been a critical contributor to political instability.

Social media platforms profit from amplifying the most engaging topics; as such, extreme, harmful, and manipulative political content spreads more rapidly and widely than ever before

(Yankoski et al., 2021). According to the United Nations, social media plays a crucial role in disseminating hate speech and incitement to violence which has real-world consequences³².

Hate speech on social media platforms does not remain online. Research shows that social media can propagate violent crimes by functioning as a platform for persuasion and collective mobilisation (Müller and Schwarz, 2021). A study of the relationship between anti-refugee hate speech on Facebook and hate crimes against refugees in Germany showed that exposure to extremist rhetoric on Facebook persuades potential perpetrators to believe that refugees are dangerous, leading to increased hate crimes against refugees (Müller and Schwarz, 2021). This research suggests that social media platforms not only facilitate the spreading of hateful ideas but also motivate real-life violence, as was seen in Myanmar.

In Myanmar, the raping, killing, and attacks on the Rohingya Muslim minority group in 2016 were partly driven by misinformation, hate speech and anti-Rohingya propaganda, which widely and rapidly spread on Facebook (Herrick, 2022). Many people in the country use Facebook as a primary news source because Facebook partnered with a local telecom company to provide subscribers with data-free access to internet services. This caused a rapid increase in the number of Facebook users in Myanmar (Mahtani, 2018). During the genocide, the platform facilitated spreading falsehoods; influential figures and top government officials used Facebook to recirculate misinformation and slurs against the Rohingya. Despite this, it was only after a UN report connected Facebook to the genocide against the Rohingya that the region became a priority for the company, which began flooding it with resources³³. Facebook has admitted its failure to prevent its platform from fuelling political division and bloodshed in Myanmar³⁴. A lawsuit has been filed against Facebook by the Rohingya refugees resettled in the United States for the platform's role in perpetuating the genocide.

³² The United Nations. "A New Era of Conflict and Violence"

<https://www.un.org/en/un75/new-era-conflict-and-violence>

³³ Amy Cheng. "Rohingya refugees sue Facebook for \$150 billion, alleging it helped perpetuate genocide in Myanmar". Washington Post, December 7, 2021. <https://www.washingtonpost.com/world/2021/12/07/facebook-rohingya-genocide-refugees-lawsuit/>

³⁴ Euan McKirdy. "We didn't do enough to prevent Myanmar violence." CNN, November 6, 2018. <https://edition.cnn.com/2018/11/06/tech/facebook-myanmar-report/index.html>

Investigations into the genocide revealed that Facebook aggressively pursued market expansion into Myanmar while neglecting adequate investment in technological funding and staffing of its operations there. In 2015, only four content moderators hired at Facebook could speak Burmese and were required to review the posts from 7.3 million Myanmar users³⁵. With limited human resources, it was impossible to examine and address all the reported content violations against the Rohingya on the platform. Additionally, the scarcity of technological resources further facilitated hate speech on Facebook. Facebook's linguistic minority users are exposed to higher levels of online hatred because Facebook's algorithm that identifies problematic content is limited to specific languages such as English, Mandarin and Spanish³⁶. Furthermore, reports show that the translation software used by the company results in inaccurate translations of Burmese script due to the software's incompatibility with some languages (Herrick, 2022).

Amidst these technological failures and the presence of ethnic tensions and other genocidal primes, Facebook did not adjust its expansion strategies or sufficiently resource the region. Even years after they expanded into the area, human rights activists and researchers warned the company that their platform was facilitating racism and hatred against Muslims in the region. Still, Facebook ignored these warnings (Herrick, 2022).

Frances Haugen, a Facebook whistle-blower stated "The raw version [of Facebook] roaming wild in most of the world does not have any of the things that make it kind of palatable in the United States. I genuinely think many lives are on the line—Myanmar and Ethiopia are like the opening chapter".³⁷

³⁵Steve Stecklow. "Why Facebook is losing the war on hate speech in Myanmar." Reuters, August 2018. <https://www.reuters.com/investigates/special-report/myanmar-facebook-hate/>

³⁶ Billy Perrigo, "Facebook Says It's Removing More Hate Speech Than Ever Before. But There's a Catch," Time, November 27, 2019, <https://time.com/5739688/facebook-hate-speech-languages/>.

³⁷ Mackintosh, Eliza. "Facebook Knew it was Being Used to Incite Violence in Ethiopia. It Did Little to Stop the Spread, Documents Show." CNN Business. October 25, 2021. <https://edition.cnn.com/2021/10/25/business/ethiopia-violence-facebook-papers-cmd-intl/index.html>

Indeed, the patterns in Myanmar are recognisable in various countries such as India, Sri Lanka, Ethiopia, and Sudan. Social media does not cause hate. However, platforms such as Facebook move into new countries without understanding the existing socio-political context. The platforms then facilitate the spread of hate speech which amplifies tensions, causes unrest, and serves as a trigger for mass violence (Herrick, 2022).

These shortcomings also reflect how social media companies, particularly Facebook, neglect non-European and American users. Facebook whistle-blower Sophie Zhang revealed that Facebook prioritises the U.S and Western European online threats, countries that pose the most public relations risk to the company. Facebook's vice-President of Integrity, Guy Rosen, believes that Facebook needs to focus its limited human resources on Western democracies, stating, "I wish resources were unlimited." At the time, the company was about to report annual operating profits of \$23.9 billion on \$70.7 billion in revenue and had \$54.86 billion in cash³⁸.

Similarly, a study shows that Facebook exposes users to misinformation and hate speech targeted at Muslims in India³⁹. Despite India constituting Facebook's largest market with 416 million regular users⁴⁰, the platform does not dedicate a reasonable amount of resources to the region. Company internal documents reveal that Facebook allocates more than 80% of its budget to fight misinformation in the United States, despite U.S users constituting only 10% of Facebook's global users³².

In many countries such as Myanmar, Mexico, and Ghana⁴¹, Facebook is the primary way users experience the internet. By neglecting to devote resources that provide vital guardrails to

³⁸ Julia Carrie Wong. "How Facebook let fake engagement distort global politics: a whistle blower's account" The Guardian. April 2021. <https://www.theguardian.com/technology/2021/apr/12/facebook-fake-engagement-whistleblower-sophie-zhang>

³⁹ Frenkel, Sheera and Davey Alba. "In India, Facebook Grapples with an Amplified Version of Its Problems." The New York Times, November 9, 2021. <https://www.nytimes.com/2021/10/23/technology/facebook-india-misinformation.html?referringSource=articleShare>.

⁴⁰World Population Review. Facebook Users by Country. <https://worldpopulationreview.com/country-rankings/facebook-users-by-country>

⁴¹Cat Zakrzewski, Gerrit De Vynck, Niha Masih and Shibani Mahtani. "How Facebook Neglected the Rest of the World" The Washington Post. October 24, 2021. <https://www.washingtonpost.com/technology/2021/10/24/india-facebook-misinformation-hate-speech/>

protect most of the global population, Facebook and other social media companies remain complicit in facilitating tragedies and endangering lives globally.

Additionally, authoritarian regimes also use social media to manipulate public discourse. A 2019 report from Oxford University discovered evidence of organised social media manipulation campaigns in 70 different countries created to influence public opinion on various issues and create political chaos (Bradshaw and Howard, 2019). The report showed that social media facilitates the spread of political propaganda, suppresses freedom of speech, and pollutes the digital information system.

For instance, the Vietnamese government threatened to shut down Facebook in the country if the company did not cooperate in censoring anti-government political content⁴². Vietnam is one of Facebook's most important regional markets, earning the company more than \$1 billion in annual revenue; as such, protecting Vietnamese users' right to free speech would have come with a high cost to the social network. Mark Zuckerberg, CEO of Facebook, personally agreed to the censorship of anti-government posts and blocking of several Vietnamese users' accounts¹¹. According to a statement provided to the Washington Post, Facebook stated that the choice to censor its users in Vietnam is justified "to ensure Facebook's services remain available for millions of people who rely on them every day." This choice to censor users, is a clear depiction of Facebook's priorities.

Conclusion

Social media is heralded for enabling coordinated civic engagement and as a platform for users to educate and express themselves. The prominence of social media underscores users' trust in these platforms. However, the ability of social media platforms to segment audiences and transmit information quickly and without regulation has facilitated the undermining of democracies globally. Bad actors use social media to manipulate public discourse in both democratic and dictatorial societies. To address these challenges, a greater understanding of the impact of technologies is needed. Additionally, there needs to be stronger regulations, rules,

⁴² Elizabeth Dwoskin, Tory Newmyer and Shibani Mahtani. "The case against Mark Zuckerberg: Insiders say Facebook's CEO chose growth over safety." Washington Post, October 25, 2021.

<https://www.washingtonpost.com/technology/2021/10/25/mark-zuckerberg-facebook-whistleblower/>

and norms for using social media, big data, and new information technologies, particularly during elections.

Chapter 5: Impact of Social Media on Individual Well-being

Within a relatively brief period, social media has revolutionised how humans interact, leaving many to wonder what implications this revolution has had on individual lives. An explosion of research has examined the impact of social media, and over time, a consistent thread has emerged. Experimental studies and research conducted by social media companies themselves show that leveraging human vulnerabilities through social media applications creates alarming and far-reaching harm to individual well-being. This chapter's objective is to investigate how social media impacts its users on a personal level, assessing the research done on the relationship between social media and variables such as mental health and well-being.

Attention and Cognition

The human brain has evolved in many brilliant ways but still has patterns which make it vulnerable to exploitation (Haselton et al., 2015). The design of social media platforms aims to keep its users engaged and influence their behaviour by taking advantage of these vulnerabilities. Thus, affecting both the attention and cognition capabilities of social media users.

Because attention is a limited resource, the salience network of the brain determines what is essential by alerting individuals to threats and opportunities in their environment (Uddin, 2016). Design features from social media applications such as notifications that utilise vibrations, red dots, flashing lights and banners constantly trigger the salience network leading to attention interruptions and disruption of focus. Social media platforms are also essential drivers of smartphone usage (Sha et al., 2019). Research conducted by Ward et al. (2017) shows that the presence of a smartphone, even when placed face down and turned off, drains attention. This experimental study of 548 adults showed that when participants' phones were turned off but on their desks instead of being in another room, it reduced working memory and the ability to solve new problems. The researchers highlight that smartphones act as "high-priority stimuli," unconsciously draining significant attentional resources even when users consciously ignore them.

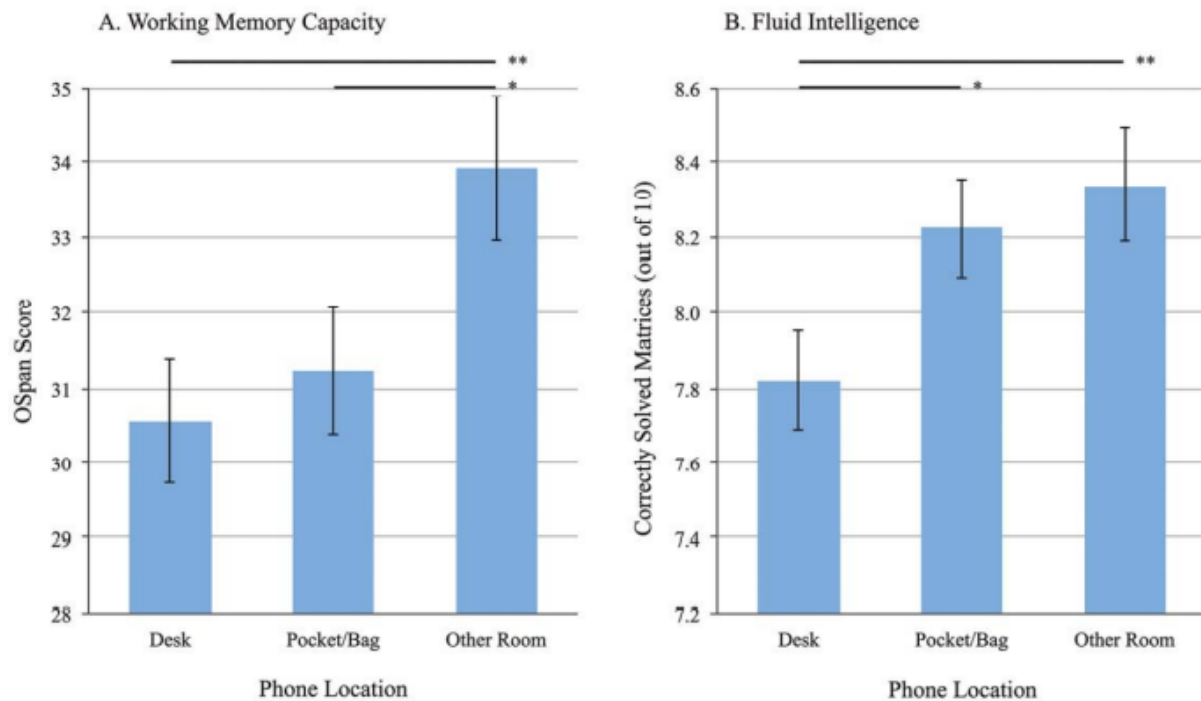


Figure 7. The Effect of randomly assigned phone location conditions on available Working Memory Capacity and Fluid Intelligence. (Source: Ward et al., 2017)

In Figure 7. Participants in the “desk” condition displayed the lowest available cognitive capacity, while those in the “other room” condition displayed the highest available cognitive capacity. Error bars represent standard errors of the means. The asterisks indicate significant differences between conditions, with significance levels of $*p < .05$ and $**p < .01$.

Similar to this study, Hadar et al. (2017) also found that heavy smartphone usage was associated with impaired attention. The brain scans from this study show that heavy smartphone users have significantly reduced neural activity in their right prefrontal cortex, a condition also seen in ADHD and linked to severe behavioural abnormalities such as impulsivity and poor attention.

Mounting evidence shows that the effects of social media on the brain are even more detrimental for young people. This is because the prefrontal cortex, which helps process and direct attention based on goals, is still developing. Unrestrained exposure to digital technology, like social media platforms, can have serious long-term consequences for development. Research suggests that screen-based media use poses neurobiological risks in children (Paulus et al., 2019, Horowitz-Kraus and Hutton, 2018, Zivan et al., 2019). Recent work showed that

pre-schoolers who use screen-based media for more than 1 hour each day have significantly less development in core brain regions involved in language and literacy (Hutton et al., 2020). Brain scans from this study indicate that the more time spent on screens, the lower the child's language skills, and the less structural integrity in crucial brain areas responsible for language.

Similarly, a longitudinal study that tracked 2441 children from the age of 2 years to 5 years old (Madigan et al., 2019) found that children with higher levels of screen time showed more significant delays in development across a range of essential measures, including language, problem-solving, and social interaction. Results from the study show a significant link between the screen time level and the specific developmental delay level 12 to 14 months later. The data from this study illustrates that exposure to excessive screen time on applications like YouTube during these early years can affect children's ability to develop optimally.

Furthermore, research on adolescents' use of social media reveals that media multitasking, such as jumping from one social media platform to another or utilising social media while simultaneously watching tv, significantly links to later levels of attentional difficulties. The study found that the degree to which teenagers multitasked was a significant predictor of attentional problems three months later, highlighting the potential impact of distracting digital environments on young teens' development. These findings support previous literature (Carrier et al., 2015, Ralph et al., 2014, Wallis, 2010), indicating that frequently engaging in multiple social media platforms for younger adolescents may lead to or worsen attention problems over time.

The effect of social media technology on younger people's attention is not only through their direct interaction with it but also through their parents' use of social media. Excessive use of social media has effects on parenting. It decreases the level of parental engagement (Leung and Lee, 2005, Kushlev and Dunn, 2019), makes children more likely to be at risk of injury (Cosottile and Damashek, 2022), and causes parental distraction (Wallis, 2010). Parents who are distracted raise children who are distracted because they teach their children the same attentional patterns (Alter, 2017). Research using head-mounted cameras shows that infants instinctively follow their parents' eyes (Yu and Smith, 2016). According to the researchers, caregivers, or parents whose eyes wander a lot or appear distracted can negatively impact an infant's developing attention span during a vital stage of their growth. An infant's ability to

sustain attention indicates later success in their cognitive development in critical areas such as problem-solving and language acquisition (Yu and Smith, 2016).

The research, therefore, provides evidence that the technology employed by social media companies may have long-term consequences for younger peoples' cognitive development and is impeding individuals' ability to focus, solve problems and be present with each other.

Mental Health and Well-being

Research conducted to study the relationship between social media use and mental health and well-being have discrepant findings. Well-being in relation to social media literature refers to how happy people are and how satisfied they are with their lives (Kross et al., 2021), while mental health is predominantly measured using constructs such as anxiety, depression, suicidal ideation and suicidal attempts.

The first group of research that examines the association between social media use and mental health or well-being utilised self-reporting, cross-sectional methods. Some of these studies revealed a positive relationship between social media usage and subjective well-being (Pittman and Reich, 2016, Apaolaza et al., 2013). Other studies reported the opposite (Leodoro, 2014, Verduyn et al., 2017). Some researchers argue that social media usage and subjective well-being have a more nuanced relationship. For instance, Rae and Lonborg (2015) found that Facebook users who utilised the platform to maintain existing relationships had a positive association between Facebook usage and subjective well-being. In contrast, Facebook usage was negatively associated with subjective well-being among those who utilised the platform to create new connections. This implies that the purpose of use may be a mediator on the impact of social media on well-being.

The second group of studies used longitudinal methods and experimental sampling for a more rigorous methodology. The outcomes, however, still produced mixed results. Some studies reported a relationship in which increased social media use was associated with a decline in well-being over time (Shakya and Christakis, 2017, Boers et al., 2019, Viner et al., 2019). In comparison, other researchers reported no substantive relationship between the two variables (Coyne et al., 2020, Jensen et al., 2019, Ferguson et al., 2014) or a relationship moderated by individual differences such as gender, hours of usage or social network size (Kelly et al., 2018, Twenge and Campbell, 2019, Przybylski and Weinstein, 2017). Interestingly, Twenge and

Campbell (2019) found that digital media users who spent less than an hour a day compared to non or moderate users had the highest levels of well-being.

Drawing conclusions from the above groups of literature is challenging for two reasons. Firstly, there are inconsistencies in how the different studies assess and operationalise the variables of well-being and social media use (Kross et al., 2021). The second challenge arises because of measures used to determine the variables. Recent research indicates that many of these studies use self-reporting measures of social media usage with questionable validity (Kross et al., 2021). For instance, a study by Ernala et al. (2020) found that participants' actual Facebook usage correlated very weakly (r between 0.23 and 0.42) with their reported measures from ten different surveys.

The last group of research utilised experimental designs. These studies used experiments to draw causal inferences on the effect of social media usage on well-being. This experimental research revealed a negative impact of social media use on well-being measures such as depression (Mosquera et al., 2020), life satisfaction (Allcott et al., 2020), mood (Tromholt, 2016) and overall subjective wellbeing (Kross et al., 2021). Although the total number of these studies remains small, their results depict an emerging pattern linking social media use to a small but significant decline in well-being.

These studies show that the impact of social media on wellbeing is nuanced and affected by various factors such as the amount of usage, the type of usage and the different psychological mechanisms that underpin social media use (Kross et al., 2021). However, there are distinct characteristics of social media that influence well-being, one of which is its tendency to be addictive.

A study by He et al. (2017) found that social media addiction had similar impacts on the brain anatomy as addictions like gambling and substance addictions. The MRI brain scans of social media users in this study demonstrated a significant reduction in grey matter in the amygdala correlated with their level of addiction to social media, particularly Facebook. The amygdala's decline in grey matter volumes has played a vital role in triggering impulsive behaviour mainly associated with addiction (Bechara, 2005). Social Networking Site addiction is a subsection of internet or technology addictions (Karaiskos et al., 2010). It refers to a user's maladaptive psychological dependency on a social networking site, such that it infringes on normal

functioning and produces conventional behavioural addiction symptoms, for instance, withdrawal, relapse, conflict and mood modification (Turel and Serenko, 2012).

Other factors on social media that influence well-being include antisocial behaviours such as cyberbullying, trolling, sexual solicitation, and harassment. Social media provides a platform for antisocial behaviour, negatively impacting other users' well-being (Kowalski et al., 2014), especially young people.

According to Giumetti and Kowalski (2022), cyberbullying primarily occurs on social media. Smith et al. (2008) define cyberbullying as "an aggressive, intentional act carried out by a group or individual, using electronic forms of contact, repeatedly and over time against a victim who cannot easily defend him or herself". Cyberbullying can occur to individuals of any age. Additionally, cyberbullying is reportedly more afflicting than "traditional bullying", potentially because the victim is aware that it is taking place in front of a much larger public audience (Van Geel et al., 2014).

Certain features of social media promote this behaviour; for instance, unlike during in-person interactions, social media interactions strip away cues that activate empathic responses (Lieberman and Schroeder, 2020, Lapidot-Lefler and Barak, 2012). Additionally, popularity on social media, measured through the number of followers or likes an individual has, validates identities as a status symbol, particularly among adolescents. (Giumetti and Kowalski, 2022). Social media users may adopt behaviours that victimise others in working towards popularity. For instance, a longitudinal study of young people in Singapore found that social media use facilitated the development of positive attitudes toward cyberbullying, which increased the likelihood of cyberbullying perpetration (Barlett et al., 2018).

Researchers have identified various adverse well-being outcomes of cyberbullying through social media among adolescents. Some of which include physical health complaints such as headaches and dizziness (Barlett et al., 2018), increased depressive symptoms and anxiety (Cao et al., 2019), lower life satisfaction (Viner et al., 2019), suicidal ideation, and suicidal attempts (Sampasa-Kanyinga and Hamilton, 2015). Employee samples found similar results, including increased psychological stress and work exhaustion (Oksanen et al., 2020).

Conclusion

The growing ubiquity of social media calls for a greater understanding of its impact on individual lives. An increasing body of research shows that excessive social media use negatively impacts users' attention, cognition, and well-being. This is particularly true amongst adolescents and developing children. The adverse effects provided by the research could lead to long-term consequences if not addressed.

Chapter 6: Regulation

The harms social media companies create and contribute to produce a multi-layered problem in which ethics, policy and business intertwine. Social media companies' lack of action to address the harms they cause is often not a lack of awareness but a deliberate choice to prioritise their bottom lines⁴³. Currently, minimal legally binding regulatory instruments act as a guardrail on human rights violations in cyberspace. As such social media companies often are not held accountable in a meaningful way as they do not face any significant consequences when they inflict harm on society. The objective of this chapter is to assess current regulations to illuminate viable solutions to address and mitigate the harms caused by social media platforms.

Regulatory Framework

Any effective regulation will require efforts from government regulators, social media companies and social media users. Addressing the challenges created and proliferated by social media will require a global stand and systematic reform.

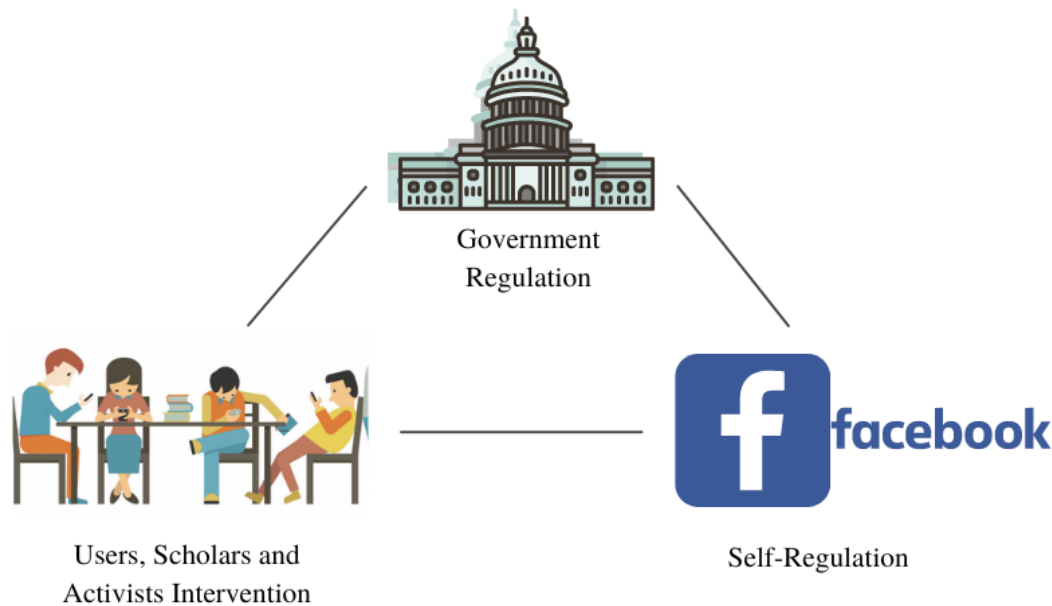


Figure 8. Regulatory framework involving government regulation, self-regulation and intervention from social media users, scholars, and activists.

⁴³ Cristiano Lima. "A whistleblower's power: Key takeaways from the Facebook Papers." Washington Post, October 26, 2021.

<https://www.washingtonpost.com/technology/2021/10/25/what-are-the-facebook-papers/>

Government Intervention

Regulators are grappling with simultaneously protecting their citizens by curbing social media companies' power while providing regulatory solutions that give freedom to corporations and promote competition, innovation, and openness⁴⁴.

Countries around the world have varying laws on social media. For instance, some states, such as China, opt for complete restriction. China has blocked access to all foreign media platforms⁴⁵. This extensive censorship system, also known as the "Great Firewall of China," means that without a VPN, Chinese citizens cannot access social media platforms like Facebook and Twitter. The government instituted this ban after unrest in the Xinjiang region in 2009 and aimed to sever communication among independence activists⁴⁶. The Chinese government subsidises the widely used alternative media application, WeChat, and the state requires the company to share users' data with it. While blocking social media platforms eliminates the harm they cause, substituting them with a centralised platform beholden to the state, does not constitute a desirable solution to the challenges posed by social media platforms.

On the opposite end of the spectrum lies the United States, which previously observed a free market approach toward digital regulation⁴⁷. The country's First Amendment right to freedom of expression creates political scrutiny on regulatory policies for social media. Additionally, Section 230 of the Communications Decency Act, initially passed in 1996, affords internet platform companies immunity from liability for almost all forms of objectionable or even

⁴⁴ Tony Raval. "Regulating Social Media Companies." Forbes. June 10, 2019.

<https://www.forbes.com/sites/forbestechcouncil/2019/06/10/regulating-social-media-companies/?sh=632cfb8162b9>

⁴⁵ Eloise Barry. "These Are the Countries Where Twitter, Facebook and TikTok Are Banned." Time. Jan 18, 2022

<https://time.com/6139988/countries-where-twitter-facebook-tiktok-banned/>

⁴⁶ Kristina Zucchi. Why Facebook is banned in China. Investopedia, December 31, 2021.

<https://www.investopedia.com/articles/investing/042915/why-facebook-banned-china.asp>

⁴⁷ Jamie Susskind. Social Media Regulation. Time, July 22, 2022. <https://time.com/6199565/regulate-social-media-platform-reduce-risks/>

illegal user-generated content on their platforms⁴⁸. The Communications Decency Act states that “no provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider.” This law ensured that platforms provided a diverse forum for public discourse that relied on voluntary self-moderation. However, in practice, the law has meant that misinformation, hate speech, and other harmful content can spread on social media. At the same time, the law protects companies from taking accountability and leaves them within their rights to keep the content available on their platforms.

More recently, there has been a shift in sentiment in the U.S towards increased regulation of Big Tech, including social media platforms. One of the proposals includes a bipartisan bill that adapts Section 230 by repealing the blanket liability that shielded social media platforms⁴⁹. Additionally, legislators have targeted large technology companies, including Facebook, for violating antitrust laws. The Federal Trade Commission sued Facebook in December 2020 for monopolising the social networking market through anticompetitive practices (Federal Trade Commission, 2020). The FTC argues that Facebook maintains its illegal monopoly through a buy or bury acquisition strategy, as evidenced by the company’s takeover of WhatsApp and Instagram. Similarly, the House of Representatives has proposed bills to curb the monopolistic power of the Big Tech companies⁵⁰. These policies challenge companies to acquire potential rivals, as Facebook did with Instagram.

Additionally, there has been increasing interest in addressing data privacy concerns. The state of California passed a consumer privacy law that many describe as a landmark policy because

⁴⁸ Dipayan Ghosh. What Will Tech Regulation Look Like in the Biden Era? Harvard Business Review, December 2020. <https://hbr.org/2020/12/what-will-tech-regulation-look-like-in-the-biden-era>

⁴⁹ The PACT Act. <https://www.schatz.senate.gov/news/press-releases/schatz-thune-introduce-new-legislation-to-update-section-230-strengthen-rules-transparency-on-online-content-moderation-hold-internet-companies-accountable-for-moderation-practices>

⁵⁰ Cecilia Kang. “Lawmakers, Taking Aim at Big Tech, Push Sweeping Overhaul of Antitrust,” The New York Times, June 11, 2021. <https://www.nytimes.com/2021/06/11/technology/big-tech-antitrust-bills.html#:~:text=WASHINGTON%E2%80%94House%20lawmakers%20on%20Friday,to%20monopoly%20laws%20in%20decades>

it constitutes the most rigorous data protection in the United States⁵¹. This law provides Californian residents with new protections. It includes the right to be informed on what type of personal data companies have collected and why it was collected. It also includes the right to request the deletion of personal information or opt-out of the sale of personal information, and the right to access the personal information in a “readily useable format” that enables its transfer to third parties without hindrance. These policies challenge social media companies because they threaten their business model. Companies such as Facebook and Twitter generate their income from advertising. Thus, these new protections afforded to individual consumers could significantly cut their profits.

While these measures indicate a step toward protecting social media users, they require more specific compliance guidelines to have sufficient impact. For instance, decreasing the amount of market power Facebook has could enable the development of new social media platforms with more robust content standards, thus moving users away from Facebook and curbing the scope and intensity of its impact. However, this outcome is far from certain, given that Facebook has a global reach and over half of the world’s Internet users regularly access the platform. Additionally, the California consumer privacy legislation has a few implications. Firstly, the law technically only applies to Californian residents, which means companies could institute a patchwork data regime that treats Californians differently than everyone else. Furthermore, this law proposes an opt-out system which means companies could use systems in which users who opt out of sharing their data would pay a premium. Companies would thus charge people for the ‘privilege’ of keeping their privacy⁵² which minimises the impact this law could have had.

Compared to the Chinese and American approaches, the European Union has proposed legislation that could substantially address the harms posed by social media platforms. The legislation offers comprehensive policies that not only require the removal of harmful content online but proactively prevent such content to be published.

⁵¹ Dipayan Ghosh. What you need to know about California’s new data privacy Law. Harvard Business Review. July 11, 2018. <https://hbr.org/2018/07/what-you-need-to-know-about-californias-new-data-privacy-law>

⁵² Jonathan Greig. “California voters back new data privacy law beefing up CCPA.” Tech Republic. <https://www.techrepublic.com/article/california-voters-back-new-data-privacy-law-beefing-up-ccpa/>

The Digital Services Act, approved by the European Parliament in January 2022, imposes policies that require external auditing of social media companies, creating stronger public oversight of the platforms⁵³. This regulation aims to make the online space safer for users by requiring social media platforms to disclose the number of content moderators they employ and what languages they speak. These policies seek greater transparency in the algorithms used by social media platforms and protect users by prohibiting companies from using deceiving or persuasive techniques to influence their behaviour. Additionally, the European Union can punish non-compliance as companies violating the Digital Services Act could face fines of up to 10% of their global revenue.

In addition to the Digital Service Act, the E.U has also proposed the Digital Markets Act to ensure fair and open competition in the digital marketplace⁵⁴. The Digital Markets Act introduces policies to curb the monopolistic power enjoyed by large technology companies like Facebook. The Act ensures that other innovators and start-ups can compete in the online platform environment without exposure to unfair terms and conditions that limit their development. Non-compliance could lead to fines of up to 10% of the company's worldwide annual turnover or up to 20% if repeated infringements occur.

The European Commission, rather than individual European states, would oversee enforcing the Digital Services Act and the Digital Markets Act. This technicality ensures the protection of users in nations such as Ireland, which have lax policies and house the European headquarters of many large technology firms. If approved by all EU Member States, the Digital Services Act and Digital Markets Act would represent the world's most comprehensive action against the harms presented by social media platforms (Herrick, 2022).

⁵³ European Parliament. "Digital Services Act: Regulating Platforms for a Safer Online Space for Users." News release, January 20, 2022, <https://www.europarl.europa.eu/news/en/press-room/20220114IPR21017/digital-services-act-regulating-platforms-for-a-safer-online-space-for-users>.

⁵⁴ European Commission, "The Digital Markets Act: Ensuring Fair and Open Digital Markets," https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/digital-markets-act-ensuring-fair-and-open-digital-markets_en

Similarly, proposed legislation aims to make the U.K. “the safest place in the world to be online”⁵⁵. The proposed bill seeks to address a range of harms, including inciting violence and violent content, encouraging suicide, disinformation, cyberbullying, and children accessing inappropriate material. The bill aims to make online companies responsible for their platforms. The Online Safety Bill is currently on pause, and some of the provisions in the bill date back five years and have not made it into law⁵⁶.

Globally, lawmakers are grappling with the imperative to address the harms facilitated by social media platforms. However, they confront a challenge as technology evolves at a pace that far outstrips the ability of legislation and bureaucracy to keep up. Social media algorithms undergo constant changes, often producing unanticipated effects and consequences that lawmakers may struggle to foresee and regulate effectively. The process of drafting impactful legislation can be time-consuming, marked by extensive deliberations and consultations, and it may ultimately culminate in legislation that is outdated or obsolete by the time it is ratified, inadvertently creating even more significant regulatory gaps (Herrick, 2022).

Moreover, a complication arises from the inherent geographical limitations of government legislation. Legal frameworks developed by sovereign entities typically operate within their respective jurisdictions, and laws passed by one country do not have universal applicability. In contrast, one of social media's primary value propositions lies in its unparalleled capacity to connect individuals and communities across international boundaries. If each nation were to assert its jurisdiction and impose specific laws governing the conduct of social media platforms, it would inevitably lead to a fragmented and disjointed regulatory landscape. This scenario risks producing a patchwork of regulations that sporadically address the harms these platforms facilitate, as each country grapples with unique challenges, priorities, and legal interpretations.

⁵⁵ UK Government. “The UK to introduce world first online safety laws.” April 8, 2019
<https://www.gov.uk/government/news/uk-to-introduce-world-first-online-safety-laws>

⁵⁶ Natasha Lomas. “UK’s Online Safety Bill on pause pending new PM.” TechCrunch. July 14, 2022.
<https://techcrunch.com/2022/07/14/online-safety-bill-pause/>

The delicate balancing act that governments face when seeking to regulate social media channels is highlighted by this regulatory landscape. On one hand, there is a pressing need to protect users from potential harms associated with social media content, such as misinformation, hate speech, and cyberbullying. On the other hand, governments must cautiously avoid stifling the innovation and free expression that social media platforms have fostered. The challenge lies in crafting regulations that strike a balance between safeguarding users' well-being and preserving the fundamental principles of freedom of speech and global connectivity.

Legislators should aim to develop regulatory frameworks that are adaptive, harmonised, and capable of addressing global challenges while respecting the unique needs and values of different nations. Achieving this equilibrium is an ongoing and complex process, one that requires continuous dialogue, adaptation, and a commitment to upholding both the rights and responsibilities that come with the immense influence of social media in the digital age.

Self-Regulation: Corporate Responsibility

While governments can play a vital role in reducing the harms facilitated by social media platforms, they are limited to their jurisdictions. Furthermore, national governments are faced with so much red tape that legislation often lacks the speed to keep up with technological progress. Platforms are better placed than regulators to understand the workings of their systems. As such, self-regulation is essential to avoid the potential tragedy of the digital commons.

For social media companies, self-regulation includes the regulatory policies that determine the company code of conduct, corporate social responsibility initiatives and terms of service that users of the platform must adhere to. Social media companies have sought to define themselves as merely objective platforms and not publishers. However, by moderating and removing content from their platforms, these companies actively take control and influence their users' experiences. Their algorithms filter content and determine what users see or do not see. They

set their own rules about what may be said and who may say it. Their principles, biases, and philosophies influence these policies. As such, they already regulate their platforms⁵⁷.

Most importantly, social media companies have strong powers of enforcement. They can approve or ban speakers and ideas on their platforms. In January 2021, Facebook, Twitter, and YouTube banned then-President Trump from their platforms, citing their rules against inciting violence. Research shows that the ban substantially decreased his online reach, with his statements gaining only a tiny fraction of his engagement before his suspension⁵⁸. While there is much debate⁵⁹ on whether the de-platforming of political leaders constitutes censorship, what remains clear is that large unelected social media companies yield the power to determine acceptable content for public discourse not just in a single country but across the world.

With great power comes great responsibility; unfortunately, social media companies have often shirked the latter. They instead opt to make decisions about acceptable content based on what yields the most significant profit irrespective of that content's impact on its users. Social media platforms' design decisions determine what users see. Their algorithms amplify divisive content that they later face pressure to remove. For social media companies to self-regulate meaningfully, they must make internal changes.

Firstly, social media platforms can reform the underlying design of their applications to mitigate the amplification of harmful content such as misinformation and hate speech. Social media algorithms customise content based on user preferences to maximise user engagement. Individuals naturally have a psychological negativity bias, which amplifies controversial and inflammatory content, creating vicious cycles of abuse and outrage on social media. Adjusting the algorithms and design choices on the platforms could address this challenge. For instance, Facebook scaled back its News Feed algorithm to reduce the existing racial and ethnic tensions

⁵⁷ Jamie Susskind. Social Media Regulation. Time, July 22, 2022. <https://time.com/6199565/regulate-social-media-platform-reduce-risks/>

⁵⁸ Jamey Alba. "What happened when Trump was Banned on Social Media." The New York Times, June 7, 2021. <https://www.nytimes.com/interactive/2021/06/07/technology/trump-social-media-ban.html>

⁵⁹ Will Oremus. "Tech giants banned Trump. But did they censor him?" The Washington Post. <https://www.washingtonpost.com/technology/2022/01/07/trump-facebook-ban-censorship/>

in volatile places like Myanmar⁶⁰. While their researchers suggested extending that decision to the rest of the platform, the company decided against it. The company's research team also found that slowing down the platform's speed encourages users to make better decisions about what they amplify. The study shows that making a design change, such as removing the reshare button, resulted in considerable gains in mitigating harm. However, Facebook would never accept such a change to their design. Reforming social media applications design can have impactful benefits for social media companies' users and society as a collective.

Additionally, not only do social media companies need to have a better understanding of the socio-political contexts of the countries they expand into, but they also need to allocate adequate resources to their operations in new markets. This measure would prevent social media platforms from amplifying existing tensions leading to tragic violence, as seen in countries like Myanmar and Ethiopia. According to leaked Facebook documents⁶¹, the company only allocates 16% of its budget to address misinformation to the "Rest of the World" category, including countries such as France, Italy, and India. 84% of the company's budget is allocated to the U.S despite U.S users only making up less than 10% of the company's daily users. This prioritisation of its Euro-American users while disregarding the majority of its global users led to Facebook's contribution to one of the worst humanitarian crises of the 21st century (Herrick, 2022). Adequate self-regulation will require social media companies to equitably invest their resources in the markets they operate.

This raises the question of motivation: Why would social media platforms take steps to reduce harm? One compelling reason is the influence of public pressure. As users become increasingly aware of the potential negative impacts of social media, they may demand responsible behaviour from these platforms. Instances of public backlash, user protests, and calls for boycotts can exert significant pressure on social media companies, compelling them to engage in self-regulation to safeguard their user base and uphold their reputation. A notable illustration of this dynamic is the potential for a massive public boycott, inclusive of major corporations,

⁶⁰ Jeff Horowitz, Ryan Knutson and Keach Hagey. The Facebook Files, Part 4: The Outrage Algorithm, podcast audio, September 18, 2021, <https://www.wsj.com/podcasts/the-journal/the-facebookfiles-part-4-the-outrage-algorithm/e619fbb7-43b0-485b-877f-18a98ffa773f>.

⁶¹ Frenkel, Sheera and Davey Alba. "In India, Facebook Grapples with an Amplified Version of Its Problems." The New York Times, November 9, 2021. <https://www.nytimes.com/2021/10/23/technology/facebook-india-misinformation.html?referringSource=articleShare>.

which could rapidly erode social media platforms' advertising revenue. Such financial repercussions would provide a strong incentive for these platforms to revise their operational processes.

For example, in a demonstration of corporate activism, prominent companies such as Unilever, North Face, and Ben & Jerry's joined the 'stop hate for profit' campaign against Facebook. A study on the campaign's impact revealed that the boycott had a detrimental effect on Facebook's stock price (Villagra, Monfort, & Méndez-Suárez, 2021). The announcement of these major companies pausing their advertising efforts significantly disrupted the firm's expected cash flows, triggering a cascading effect that exacerbated the situation (Villagra, Monfort, & Méndez-Suárez, 2021). Such substantial financial consequences can serve as a compelling motivator for social media platforms to address issues related to harmful content and practices, aligning their interests more closely with the public's concerns about responsible behaviour.

However, it's essential to acknowledge that while the campaign did have a noteworthy financial impact on Facebook, it did not lead to radical or enduring changes in the platform's content moderation policies and practices.

Another potential catalyst for change lies in a shift in leadership. Changes in leadership within these companies can spark initiatives aimed at self-regulation. CEOs and executives who prioritize social responsibility and ethical behaviour may implement policies and practices designed to enhance the overall quality of their platforms. A change in leadership could represent an opportunity to reshape the corporate structure and prioritize human rights, even if it entails marginal setbacks in the platform's growth rate. Although unlikely, an internal transformation that redirects social media companies' focus beyond growth and revenue generation could be a viable means of prioritizing the safety and well-being of the platform's users.

Nonetheless, the threat of legal action and regulatory pressure has consistently proven to be one of the most effective drivers of change across various industries (Cusumano, Gawer, & Yoffie, 2021). Self-regulation often emerges as a response to the credible threat of government regulation. This pattern has been observed in industries such as tobacco and cigarette advertising, as well as airline reservations (Cusumano, Gawer, & Yoffie, 2021). The mere anticipation of government regulations can motivate social media companies to engage in self-regulation as a proactive measure, demonstrating their commitment to addressing issues and

circumventing the potential imposition of more stringent regulations that could hinder innovation and growth.

Conclusion

Globally, governments are grappling with how to regulate social media. Some countries like China have completely banned specific social media platforms in their country, and other states such as the U.S. and the E.U. are creating comprehensive regulatory frameworks to mitigate their harms. Government intervention does not come without its own set of challenges. However, a reasonable aim would be to reduce harm rather than seek regulatory perfection. Laws should reasonably align with incentives (Lanier, 2018). Legislation by itself may be insufficient for impactful change, social media companies must reform their policies to mitigate the problems they proliferate. In the future, governments and social media companies will need to work together to maximise the benefits of these digital platforms while minimising their harm to individuals and society.

Chapter 7: Limitations and Future Research

This review extensively examined the impact of social media on individual and societal well-being. However, the primary research evidence, as revealed through the exploration of identified themes, presented limitations. While concerning results emerged regarding the impact of social media on society, experimental research in this realm remains limited, contributing to a scarcity of knowledge concerning the long-term effects of these technologies. Similarly, the evidence regarding the impact of social media on individual well-being was inconclusive.

Methodologically, this thesis critically assessed individual studies and their outcomes. However, the diversity across these studies, spanning various disciplines such as psychology, communication, and sociology, presented challenges in comparing findings. Variations in study populations, comparisons, and outcomes resulted in contradictory findings within the

reviewed literature. The heterogeneous nature of the collected research limited the feasibility of sensitivity analysis, subgroup analysis, or meta-analyses. The rapid evolution of social media platforms and user practices further posed challenges for researchers, complicating measurement and comparison efforts. Moreover, the methodology employed in this thesis could identify associations and trends but could not establish causation or correlation due to the heterogeneous nature of the literature and the research objectives. This limitation restricts the ability to establish firm cause-and-effect relationships.

Future Research Directions

To address these limitations, future research could consider correlational or regression analyses to explore relationships between studies and investigate potential impacts of interventions on social media use. Additionally, given the rapid evolution of social media technologies and user behaviours, researchers must adopt more responsive evaluation approaches to accommodate these changes. The dynamic nature of these platforms demands evaluations that remain relevant to the current version of the tool being investigated.

Furthermore, to enhance future research, a call is made for more consistent and comprehensive definitions of social media, encompassing more recent platforms like Tiktok, Instagram threads and others. Improved and standardised definitions will ensure the inclusivity of contemporary tools in research studies.

Chapter 8: Conclusion

Social media is here to stay. Its global ubiquity and growing popularity show that it fulfils needs and desires. In some ways, social media has had a positive impact. It enables the mobilisation of individuals around prominent issues, facilitates the sharing of knowledge across borders and at zero or marginal cost connects billions of users globally. However, this starry-eyed view of social media is not where the story ends. Once hailed as a beacon of hope for a more liberated world, the pendulum on social media has now swung in the opposite direction and asks that we take a closer look.

Upon closer inspection, it is clear that social media divides people as much as it connects them. The idealistic vision of people sharing more information with others worldwide has not improved nations or global culture, enhanced mutual understanding, or strengthened democratic movements.

Yes, social media companies have generated trillions of dollars in wealth. However, they have also facilitated the distribution of fake news, manipulated digital content for political purposes, and promoted dangerous misinformation on elections, vaccines, and other public health matters. Social media companies have amplified tensions in certain regions, endangering lives and contributing to tragedies globally.

It is natural to ask if social media is the cause of society's ills or merely a mirror reflecting them. This thesis does not argue that social media is a source of all these challenges. However, social media is a powerful tool that amplifies the worst of society. Social media companies introduce their products and services into new regions without understanding or caring about sociopolitical contexts. While these platforms do not create social upheaval, they certainly exacerbate it. The underlying design of these platforms intensify negativity, extremism, and divisiveness. Ultimately, social media companies prioritise their bottom line while neglecting their users, especially those in the Global South.

The social dilemma is apparent; social media platforms can foster evil and good. Regulatory interventions will be necessary for mitigating the problems associated with social media while maximising the benefits. Policies in anti-trust and competition, privacy and data protection will be essential for protecting citizens and systematically addressing the harms proliferated by digital platforms.

Social media companies have the responsibility to be cognizant of the impact their products have on society and work towards addressing them. It will be impossible to meaningfully address these harms without the companies' internal reform. One of Facebook's founders' more famous mottos is "move fast and break things"⁶². Now that the company has faced global scrutiny over privacy, misinformation, harmful content and fostering genocide, perhaps enough things have been broken; perhaps Facebook could start working towards fixing them.

In the long term, there is hope. The community of activists, writers, policymakers, critics, and scholars concerned about these challenges continue to grow. Any effective response to the problems will require concerted recognition and awareness of the damage done by social media

⁶² Hemant Taneja. The Era of "Move Fast and Break Things" Is Over. January 2019. <https://hbr.org/2019/01/the-era-of-move-fast-and-break-things-is-over>

platforms. In your reading of this thesis, we have already made the first step in the right direction.

References

- ALLCOTT, H., BRAGHIERI, L., EICHMEYER, S. & GENTZKOW, M. 2020. The welfare effects of social media. *American Economic Review*, 110, 629-76.
- ALLCOTT, H. & GENTZKOW, M. 2017. Social media and fake news in the 2016 election. *Journal of economic perspectives*, 31, 211-36.
- ALTER, A. 2017. *Irresistible: The rise of addictive technology and the business of keeping us hooked*, Penguin.
- APAOLAZA, V., HARTMANN, P., MEDINA, E., BARRUTIA, J. M. & ECHEBARRIA, C. 2013. The relationship between socializing on the Spanish online networking site Tuenti and teenagers' subjective wellbeing: The roles of self-esteem and loneliness. *Computers in Human Behavior*, 29, 1282-1289.
- APUKE, O. D. & OMAR, B. 2021. Fake news and COVID-19: modelling the predictors of fake news sharing among social media users. *Telematics and Informatics*, 56, 101475.
- AQUINO, F., DONZELLI, G., DE FRANCO, E., PRIVITERA, G., LOPALCO, P. L. & CARDUCCI, A. 2017. The web and public confidence in MMR vaccination in Italy. *Vaccine*, 35, 4494-4498.

- AYMANN, C., FOERSTER, J. & GEORG, C.-P. 2017. Fake news in social networks. *arXiv preprint arXiv:1708.06233*.
- AZZIMONTI, M. & FERNANDES, M. 2018. Social media networks, fake news, and polarization. National Bureau of Economic Research.
- BARLETT, C. P., GENTILE, D. A., CHNG, G., LI, D. & CHAMBERLIN, K. 2018. Social media use and cyberbullying perpetration: A longitudinal analysis. *Violence and gender*, 5, 191-197.
- BAUMANN, F., LORENZ-SPREEN, P., SOKOLOV, I. M. & STARNINI, M. 2020. Modeling echo chambers and polarization dynamics in social networks. *Physical Review Letters*, 124, 048301.
- BECHARA, A. 2005. Decision making, impulse control and loss of willpower to resist drugs: a neurocognitive perspective. *Nature neuroscience*, 8, 1458-1463.
- BESSI, A., ZOLLO, F., DEL VICARIO, M., PULIGA, M., SCALA, A., CALDARELLI, G., UZZI, B. & QUATTROCIOCCI, W. 2016. Users polarization on Facebook and Youtube. *PLoS one*, 11, e0159641.
- BOERS, E., AFZALI, M. H., NEWTON, N. & CONROD, P. 2019. Association of screen time and depression in adolescence. *JAMA pediatrics*, 173, 853-859.
- BOXELL, L., GENTZKOW, M. & SHAPIRO, J. M. 2017. Greater Internet use is not associated with faster growth in political polarization among US demographic groups. *Proceedings of the National Academy of Sciences*, 114, 10612-10617.
- BOYD, D. M. & ELLISON, N. B. 2007. Social network sites: Definition, history, and scholarship. *Journal of computer-mediated Communication*, 13, 210-230.
- BRADSHAW, S. & HOWARD, P. N. 2019. The global disinformation order: 2019 global inventory of organised social media manipulation.
- BRAUN, V., & CLARKE, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2), 77-101.
- CAO, X., KHAN, A. N., ZAIGHAM, G. H. & KHAN, N. A. 2019. The stimulators of social media fatigue among students: Role of moral disengagement. *Journal of Educational Computing Research*, 57, 1083-1107.
- CARRIER, L. M., ROSEN, L. D., CHEEVER, N. A. & LIM, A. F. 2015. Causes, effects, and practicalities of everyday multitasking. *Developmental Review*, 35, 64-78.
- CARRIERI, V., MADIO, L. & PRINCIPE, F. 2019. Vaccine hesitancy and (fake) news: Quasi-experimental evidence from Italy. *Health economics*, 28, 1377-1382.
- CINELLI, M., MORALES, G. D. F., GALEAZZI, A., QUATTROCIOCCI, W. & STARNINI, M. 2021. The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, 118.
- CORBYN, Z. 2012. Facebook experiment boosts US voter turnout. *Nature*, 12.
- COSOTTILE, M. & DAMASHEK, A. 2022. Effects of cell phone use on caregiver supervision and child injury risk. *Journal of pediatric psychology*, 47, 86-93.

- COYNE, S. M., ROGERS, A. A., ZURCHER, J. D., STOCKDALE, L. & BOOTH, M. 2020. Does time spent using social media impact mental health?: An eight year longitudinal study. *Computers in Human Behavior*, 104, 106160.
- Cusumano, M. A., Gawer, A., & Yoffie, D. B. (2021). Social media companies should self-regulate. Now. *Harvard Business Review*, 15.
- DI DOMENICO, G., SIT, J., ISHIZAKA, A. & NUNAN, D. 2021. Fake news, social media and marketing: A systematic review. *Journal of Business Research*, 124, 329-341.
- ERNALA, S. K., BURKE, M., LEAVITT, A. & ELLISON, N. B. How well do people report time spent on Facebook? An evaluation of established survey questions with recommendations. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 2020. 1-14.
- FERGUSON, C. J., MUÑOZ, M. E., GARZA, A. & GALINDO, M. 2014. Concurrent and prospective analyses of peer, television and social media influences on body dissatisfaction, eating disorder symptoms and life satisfaction in adolescent girls. *Journal of youth and adolescence*, 43, 1-14.
- FERRARA, E., VAROL, O., DAVIS, C., MENCZER, F. & FLAMMINI, A. 2016. The rise of social bots. *Communications of the ACM*, 59, 96-104.
- FLORES, K. 2022. *Shifting Social Media Incentives with the Subscription Model* [Online]. Available: <https://www.openthoughtblog.com/shifting-social-media-incentives-with-the-subscription-model/> [Accessed 29th May 2022].
- GIUMETTI, G. W. & KOWALSKI, R. M. 2022. Cyberbullying via social media and well-being. *Current Opinion in Psychology*, 101314.
- GONZÁLEZ-PADILLA, D. A. & TORTOLERO-BLANCO, L. 2020. Social media influence in the COVID-19 Pandemic. *International braz j urol*, 46, 120-124.
- HASELTON, M. G., NETTLE, D. & ANDREWS, P. W. 2015. The evolution of cognitive bias. *The handbook of evolutionary psychology*, 724-746.
- HE, Q., TUREL, O. & BECHARA, A. 2017. Brain anatomy alterations associated with Social Networking Site (SNS) addiction. *Scientific reports*, 7, 1-8.
- HERRICK, K. 2022. Breaking Things: Origins and Consequences of Racialized Hate Speech on Facebook.
- HOROWITZ-KRAUS, T. & HUTTON, J. S. 2018. Brain connectivity in children is increased by the time they spend reading books and decreased by the length of exposure to screen-based media. *Acta paediatrica*, 107, 685-693.
- HOWARD, P. N., KOLLANYI, B., BRADSHAW, S. & NEUDERT, L.-M. 2018. Social media, news and political information during the US election: Was polarizing content concentrated in swing states? *arXiv preprint arXiv:1802.03573*.
- HUTTON, J. S., DUDLEY, J., HOROWITZ-KRAUS, T., DEWITT, T. & HOLLAND, S. K. 2020. Associations between screen-based media use and brain white matter integrity in preschool-aged children. *JAMA pediatrics*, 174, e193869-e193869.
- ISAAK, J. & HANNA, M. J. 2018. User data privacy: Facebook, Cambridge Analytica, and privacy protection. *Computer*, 51, 56-59.

- JAMIESON, K. H. 2020. *Cyberwar: how Russian hackers and trolls helped elect a president: what we don't, can't, and do know*, Oxford University Press.
- JENSEN, M., GEORGE, M. J., RUSSELL, M. R. & ODGERS, C. L. 2019. Young adolescents' digital technology use and mental health symptoms: Little evidence of longitudinal or daily linkages. *Clinical Psychological Science*, 7, 1416-1433.
- KARAISKOS, D., TZAVELLAS, E., BALTA, G. & PAPARRIGOPOULOS, T. 2010. P02-232-Social network addiction: a new clinical disorder? *European Psychiatry*, 25, 1-1.
- KELLY, Y., ZILANAWALA, A., BOOKER, C. & SACKER, A. 2018. Social media use and adolescent mental health: Findings from the UK Millennium Cohort Study. *EClinicalMedicine*, 6, 59-68.
- KILLICK, C., & Taylor, B. J. (2009). Professional decision making on elder abuse: Systematic narrative review. *Journal of Elder Abuse & Neglect*, 21(3), 211-238.
- KIM, A. & DENNIS, A. R. 2019. Says who? The effects of presentation format and source rating on fake news in social media. *MIS quarterly*, 43, 1025-1039.
- KOWALSKI, R. M., GIUMETTI, G. W., SCHROEDER, A. N. & LATTANNER, M. R. 2014. Bullying in the digital age: a critical review and meta-analysis of cyberbullying research among youth. *Psychological bulletin*, 140, 1073.
- KRAMER, A. D., GUILLORY, J. E. & HANCOCK, J. T. 2014. Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 111, 8788-8790.
- KROSS, E., VERDUYN, P., SHEPPES, G., COSTELLO, C. K., JONIDES, J. & YBARRA, O. 2021. Social media and well-being: Pitfalls, progress, and next steps. *Trends in Cognitive Sciences*, 25, 55-66.
- KUSHLEV, K. & DUNN, E. W. 2019. Smartphones distract parents from cultivating feelings of connection when spending time with their children. *Journal of Social and Personal Relationships*, 36, 1619-1639.
- LANIER, J. 2018. *Ten arguments for deleting your social media accounts right now*, Random House.
- LANUZA, E., NOVEJARQUE, A., MARTÍNEZ-RICÓS, J., MARTÍNEZ-HERNÁNDEZ, J., AGUSTÍN-PAVÓN, C. & MARTÍNEZ-GARCÍA, F. 2008. Sexual pheromones and the evolution of the reward system of the brain: the chemosensory function of the amygdala. *Brain Research Bulletin*, 75, 460-466.
- LAPIDOT-LEFLER, N. & BARAK, A. 2012. Effects of anonymity, invisibility, and lack of eye-contact on toxic online disinhibition. *Computers in human behavior*, 28, 434-443.
- LAZER, D. M., BAUM, M. A., BENKLER, Y., BERINSKY, A. J., GREENHILL, K. M., MENCZER, F., METZGER, M. J., NYHAN, B., PENNYCOOK, G. & ROTHSCHILD, D. 2018. The science of fake news. *Science*, 359, 1094-1096.
- LEODORO, L. 2014. Facebook use and adolescents's emotional states of depression, anxiety, and stress. *Health science journal*, 8, 0-0.

- LEUNG, L. & LEE, P. S. 2005. Multiple determinants of life quality: The roles of Internet activities, use of new media, social support, and leisure activities. *Telematics and Informatics*, 22, 161-180.
- LEWANDOWSKY, S., ECKER, U. K., SEIFERT, C. M., SCHWARZ, N. & COOK, J. 2012. Misinformation and its correction: Continued influence and successful debiasing. *Psychological science in the public interest*, 13, 106-131.
- LEWIS, R. 2018. Alternative influence: Broadcasting the reactionary right on YouTube.
- LIEBERMAN, A. & SCHROEDER, J. 2020. Two social lives: How differences between online and offline interaction influence social outcomes. *Current Opinion in Psychology*, 31, 16-21.
- LIN, L. 2017. TikTok Owner ByteDance's Annual Revenue Jumps to \$34.3 Billion. *Wall Street Journal*.
- MADIGAN, S., BROWNE, D., RACINE, N., MORI, C. & TOUGH, S. 2019. Association between screen time and children's performance on a developmental screening test. *JAMA pediatrics*, 173, 244-250.
- MAHTANI, S. 2018. UN report calls for Myanmar generals to be prosecuted for genocide, war crimes. *Washington Post*, August, 27.
- MATAKOS, A., TERZI, E. & TSAPARAS, P. 2017. Measuring and moderating opinion polarization in social networks. *Data Mining and Knowledge Discovery*, 31, 1480-1505.
- MAYFIELD, A. 2008. What is social media.
- MCFADDEN, P., Taylor, B. J., Campbell, A., & McQuilkin, J. (2012). Systematically identifying relevant research: Case study on child protection social workers' resilience. *Research on Social Work Practice*, 22(6), 626-636.
- MCCRAE, N., & PURSSELL, E. (2015). Eligibility criteria in systematic reviews published in prominent medical journals: a methodological review. *Journal of Evaluation in Clinical Practice*, 21(6), 1052-1058
- MCGONAGLE, T. 2017. "Fake news" False fears or real concerns? *Netherlands Quarterly of Human Rights*, 35, 203-209.
- MESHI, D., TAMIR, D. I. & HEEKEREN, H. R. 2015. The emerging neuroscience of social media. *Trends in cognitive sciences*, 19, 771-782.
- MITCHELL, A., GOTTFRIED, J., STOCKING, G., WALKER, M. & FEDELI, S. 2019. Many Americans say made-up news is a critical problem that needs to be fixed. *Pew Research Center*, 5, 2019.
- MOSQUERA, R., ODUNOWO, M., MCNAMARA, T., GUO, X. & PETRIE, R. 2020. The economic effects of Facebook. *Experimental Economics*, 23, 575-602.
- MÜLLER, K. & SCHWARZ, C. 2021. Fanning the flames of hate: Social media and hate crime. *Journal of the European Economic Association*, 19, 2131-2167.
- OKSANEN, A., OKSA, R., SAVELA, N., KAAKINEN, M. & ELLONEN, N. 2020. Cyberbullying victimization at work: Social media identity bubble approach. *Computers in human behavior*, 109, 106363.

- PAULUS, M. P., SQUEGLIA, L. M., BAGOT, K., JACOBUS, J., KUPLICKI, R., BRESLIN, F. J., BODURKA, J., MORRIS, A. S., THOMPSON, W. K. & BARTSCH, H. 2019. Screen media activity and brain structure in youth: evidence for diverse structural correlation networks from the ABCD study. *Neuroimage*, 185, 140-153.
- PENNYCOOK, G., CANNON, T. D. & RAND, D. G. 2018. Prior exposure increases perceived accuracy of fake news. *Journal of experimental psychology: general*, 147, 1865.
- PITTMAN, M. & REICH, B. 2016. Social media and loneliness: Why an Instagram picture may be worth more than a thousand Twitter words. *Computers in Human Behavior*, 62, 155-167.
- PRZYBYLSKI, A. K. & WEINSTEIN, N. 2017. A large-scale test of the goldilocks hypothesis: quantifying the relations between digital-screen use and the mental well-being of adolescents. *Psychological science*, 28, 204-215.
- RAE, J. R. & LONBORG, S. D. 2015. Do motivations for using Facebook moderate the association between Facebook use and psychological well-being? *Frontiers in psychology*, 6, 771.
- RALPH, B. C., THOMSON, D. R., CHEYNE, J. A. & SMILEK, D. 2014. Media multitasking and failures of attention in everyday life. *Psychological research*, 78, 661-669.
- RIBEIRO, M. H., OTTONI, R., WEST, R., ALMEIDA, V. A. & MEIRA JR, W. Auditing radicalization pathways on YouTube. Proceedings of the 2020 conference on fairness, accountability, and transparency, 2020. 131-141.
- SAMPASA-KANYINGA, H. & HAMILTON, H. 2015. Social networking sites and mental health problems in adolescents: The mediating role of cyberbullying victimization. *European psychiatry*, 30, 1021-1027.
- SHA, P., SARIYSKA, R., RIEDL, R., LACHMANN, B. & MONTAG, C. 2019. Linking internet communication and smartphone use disorder by taking a closer look at the Facebook and WhatsApp applications. *Addictive behaviors reports*, 9, 100148.
- SHAKYA, H. B. & CHRISTAKIS, N. A. 2017. Association of Facebook use with compromised well-being: A longitudinal study. *American journal of epidemiology*, 185, 203-211.
- SHAO, C., CIAMPAGLIA, G. L., VAROL, O., YANG, K.-C., FLAMMINI, A. & MENCZER, F. 2018. The spread of low-credibility content by social bots. *Nature communications*, 9, 1-9.
- SIMON, H. A. 1971. Designing organizations for an information-rich world. *Computers, communications, and the public interest*, 72, 37.
- SMITH, P. K., MAHDAVI, J., CARVALHO, M., FISHER, S., RUSSELL, S. & TIPPETT, N. 2008. Cyberbullying: Its nature and impact in secondary school pupils. *Journal of child psychology and psychiatry*, 49, 376-385.
- TAMBINI, D. 2018. Social media power and election legitimacy.
- TROMHOLT, M. 2016. The Facebook experiment: Quitting Facebook leads to higher levels of well-being. *Cyberpsychology, behavior, and social networking*, 19, 661-666.
- TUCKER, J. A., GUESS, A., BARBERÁ, P., VACCARI, C., SIEGEL, A., SANOVICH, S., STUKAL, D. & NYHAN, B. 2018. Social media, political polarization, and political disinformation: A

review of the scientific literature. *Political polarization, and political disinformation: a review of the scientific literature (March 19, 2018)*.

- TUFEKCI, Z. & WILSON, C. 2012. Social media and the decision to participate in political protest: Observations from Tahrir Square. *Journal of communication*, 62, 363-379.
- TUREL, O. & SERENKO, A. 2012. The benefits and dangers of enjoyment with social networking websites. *European Journal of Information Systems*, 21, 512-528.
- TWENGE, J. M. & CAMPBELL, W. K. 2019. Media use is linked to lower psychological well-being: Evidence from three datasets. *Psychiatric Quarterly*, 90, 311-331.
- UDDIN, L. Q. 2016. *Saliency network of the human brain*, Academic press.
- VAIDHYANATHAN, S. 2018. *Antisocial media: How Facebook disconnects us and undermines democracy*, Oxford University Press.
- VAN DUYN, E. & COLLIER, J. 2019. Priming and fake news: The effects of elite discourse on evaluations of news media. *Mass Communication and Society*, 22, 29-48.
- VAN GEEL, M., VEDDER, P. & TANILON, J. 2014. Relationship between peer victimization, cyberbullying, and suicide in children and adolescents: a meta-analysis. *JAMA pediatrics*, 168, 435-442.
- VERDUYN, P., YBARRA, O., RÉSIBOIS, M., JONIDES, J. & KROSS, E. 2017. Do social network sites enhance or undermine subjective well-being? A critical review. *Social Issues and Policy Review*, 11, 274-302.
- Villagra, N., Monfort, A., & Méndez-Suárez, M. (2021). Firm value impact of corporate activism: Facebook and the stop hate for profit campaign. *Journal of Business Research*, 137, 319-326.
- VINER, R. M., GIREESH, A., STIGLIC, N., HUDSON, L. D., GODDINGS, A.-L., WARD, J. L. & NICHOLLS, D. E. 2019. Roles of cyberbullying, sleep, and physical activity in mediating the effects of social media use on mental health and wellbeing among young people in England: a secondary analysis of longitudinal data. *The Lancet Child & Adolescent Health*, 3, 685-696.
- VOSOUGHI, S., ROY, D. & ARAL, S. 2018. The spread of true and false news online. *Science*, 359, 1146-1151.
- WALLIS, C. The impacts of media multitasking on children's learning and development: Report from a research seminar. The Joan Ganz Cooney Center at Sesame Workshop, New York, 2010.
- WARD, A. F., DUKE, K., GNEEZY, A. & BOS, M. W. 2017. Brain drain: The mere presence of one's own smartphone reduces available cognitive capacity. *Journal of the Association for Consumer Research*, 2, 140-154.
- YANKOSKI, M., THEISEN, W., VERDEJA, E. & SCHEIRER, W. J. 2021. Artificial Intelligence for Peace: An Early Warning System for Mass Violence. *Towards an International Political Economy of Artificial Intelligence*. Springer.
- YU, C. & SMITH, L. B. 2016. The social origins of sustained attention in one-year-old human infants. *Current biology*, 26, 1235-1240.
- ZHANG, X. & GHORBANI, A. A. 2020. An overview of online fake news: Characterization, detection, and discussion. *Information Processing & Management*, 57, 102025.

ZIVAN, M., BAR, S., JING, X., HUTTON, J., FARAH, R. & HOROWITZ-KRAUS, T. 2019. Screen-exposure and altered brain activation related to attention in preschool children: An EEG study. *Trends in Neuroscience and Education*, 17, 100117.