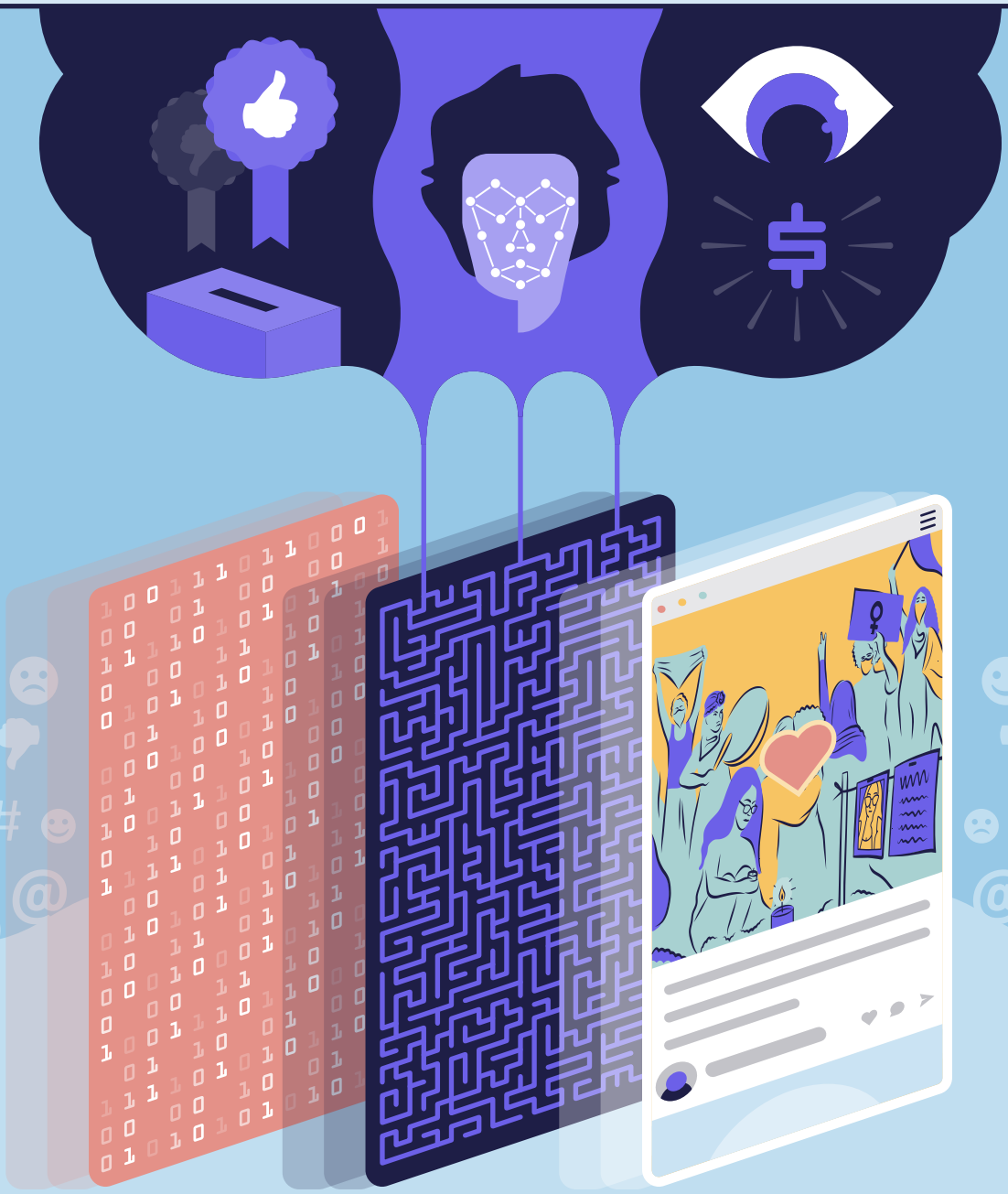
ALiGN Advancing Learning and Innovation on Gender Norms

ODI Think Change

# HIDDEN IN PLAIN SIGHT

## HOW THE INFRASTRUCTURE OF SOCIAL MEDIA SHAPES GENDER NORMS



Stephanie Diepeveen

ALiGN

## Publication information

**Published by:** Advancing Learning and Innovation on Gender Norms (ALIGN) and ODI, February 2022.


## Suggested citation and permalink

Diepeveen, S. (2022) *Hidden in plain sight: how the infrastructure of social media shapes gender norms*. ALIGN Report. London: ODI (https://www.alignplatform.org/resources/report-hidden-in-plain-sight).

## Acknowledgements

Author: Stephanie Diepeveen

Editors: Terese Jonsson

This report benefitted from the contributions and insights of many people throughout.

A particular thank you to Emilie Tant for working on the structure and for communications support, and to Rachel Marcus, Caroline Harper and Rachel George for helpful advice and reviews during the drafting process. I am also grateful to the invaluable comments provided by peer reviewers Eleanor Drage and Louise Shaxson.

Finally, thank you to Sonia Hoque who provided administrative and budgetary management, Terese Jonsson for copy-editing, Emily Subden for publications delivery, Emma Carter for proofreading, and also to Steven Dickie for typeset and design. All remaining errors are the author's own.

This study was undertaken with financial support from the Government of Canada. The views expressed are those of the authors and do not necessarily reflect the official views or policies of GAC.

## About the authors

**Stephanie Diepeveen** is a Research Fellow in ODI's Politics and Governance programme and plays a lead role in the institute's Digital Societies initiative. Stephanie is also a Research Associate in the Department of Politics and International Studies at the University of Cambridge.

Contents

# Acronyms

| | |
|---|---|
| **API** | application programming interface |
| **CDA** | Communications Decency Act (US) |
| **CEO** | Chief Executive Officer |
| **GDPR** | General Data Protection Regulation (EU) |
| **EU** | European Union |
| **LGBTQI+** | lesbian, gay, bisexual, trans, queer/questioning, intersex (+ denotes that the acronym is non-exhaustive and can also include other identities) |
| **NSFW** | Not Safe For Work |
| **UK** | United Kingdom |
| **US** | United States |

# Display items

## Boxes



## Figures



## Tables

# Glossary

**Algorithm –** A set of instructions or rules on how to deal with information. Information is inputted into algorithms as data. Algorithms process data according to precise steps in order to solve a problem or complete a task.

**App store –** A platform that allows users to search for different software and applications (apps) to download.

**Application programming interface (API) –** A software interface that allows two applications to communicate and access one another's data. They allow for interoperability (i.e. sharing of information) between websites and online services, enabling people to interact through connectivity applications (e.g. the Facebook 'Like' button) and different websites to access data (Bodle, 2011).

**Crowd work –** Task-based work, whereby self-employed workers pick up 'tasks' via a digital platform; part of the digital 'gig' economy.

**Data –** In the context of social media platforms, data is the information that is generated and recorded from users' digital activity.

**Data point –** An identifiable element in a dataset.

**Data source (first-, second- and third-party) –** Data is defined as coming from a first-, second- or third-party data source depending on where the data originates. First-party data is information collected from a company's own sources. Second-party data comes from another company, and is that company's first-party data. Third-party data comes from a company that does not have a direct relationship with the data.

**Dataset –** A collection of data.

**Filter –** Software that is designed to sort and limit access to specific content.

**Flag, flagging –** A mechanism for users to report offensive content on social media platforms.

**Gender norms –** Socio-culturally defined rules about how a person should behave and present themselves in accordance with their perceived gender.

**Heteronormativity –** Views about gender and sexuality rooted in a belief that heterosexuality and binary gender categories (man/woman) are natural and normal, with other sexualities and gender identities considered to be deviant.

**Influencer –** A popular social media user who is paid by brands or marketing firms to advertise products to their networks.

**Internet service provider –** An organisation that provides internet connections and services to individuals, groups and companies.

**Intersectionality –** A concept, originating in black feminism, which explains how different social categories interconnect to produce differing experiences of oppression and privilege for different groups of people (such as depending on their race/ethnicity, gender, class and sexuality).

**Machine learning –** Algorithms that enable a computer to learn without ongoing human input. In basic terms, a computer learns by finding patterns in data and applying these patterns.

**NSFW (Not Safe For Work) –** A content warning online to indicate sensitive content that may not be appropriate for public viewing.

**Principal component analysis –** A statistical process to reduce the number of attributes in a dataset with minimal information loss, by organising a large number of attributes into a smaller number of simpler categories.

**Protocol (for code) –** A set of rules for presenting and processing data that allows, for example, different computers to communicate.

**Search algorithm –** A set of rules or instructions (algorithms) to retrieve specific data from a wider collection of data.

**Shadowbanning –** The practice, used by some social media platforms, of blocking or limiting the visibility of a users' content without their knowledge.

**Social media platform –** This report uses Carr and Hayes' (2015: 50) definition of social media platforms as 'Internet-based channels that allow users to opportunistically interact and selectively self-present, either in real-time or asynchronously, with both broad and narrow audiences who derive value from user-generated content and the perception of interaction with others'.

**Software development kit –** A collection of software tools, for example to develop an application on a platform. This can include APIs.

*We discuss a selective list of social media platforms in this report, focusing on the platforms that have emerged as dominant in the 2010s and in existing research. This includes: social networking sites, specifically Twitter, Facebook and Tumblr; the image and video-based social media applications Instagram and TikTok; the video sharing platform YouTube; messaging applications, including WhatsApp, Signal and Telegram; and finally the social dating application Bumble.*

**Content warning –** *This report contains some quotes of sexist language that promotes violence against women (specifically in Sub-section 3.4).*

# 1 Introduction

Social media has become one of the 21st century's most powerful and era-defining innovations. It has come to permeate everyday human interactions, becoming an increasingly irreplaceable communications tool for individuals and organisations all over the world. Online platforms have infiltrated the very functioning of our personal and intimate relationships, the marketplace and the political sphere, and they have blurred the boundaries between local and global communities and driven cultural trends.

**Social media platforms** have thus evolved into valuable mediums for people to interact globally, often mediating social and personal relations, business transactions, access to entertainment and the consumption of news and public affairs stories.

Much can be said for the efficiencies and endless possibilities for connection that social media provides. At the same time, more critical questions are also being asked about how this new realm of social interaction is impacting and shaping broader social structures and relationships. For instance, there is growing attention to the way that people's use of online platforms influences gender identities and relations.

This report is one half of a two-part ALIGN research series focusing on social media and **gender norms**, which explores how gender relations shape, and are shaped by, the proliferation of online social networks. As a point of departure, this report focuses on the building blocks of social media that users do not see: the back-end infrastructures of the platforms. By examining this material, technological and institutional architecture through a gender lens, this research aims to build knowledge about the possibilities for changing patriarchal social norms in the internet age. A companion report centres on the front end of social media, unpacking the extent to which online activism presents meaningful opportunities to shift gender norms (Washington and Marcus, 2022).

While the potential for change appears to lie in people's active use of and interactions through social media, to fully conceptualise the pathways or routes to transformation, it is important to take into account how the spaces in which people interact are constructed. These structures are hidden in plain sight, yet are critical to facilitating and shaping how people express or present their gendered selves online. Social media is an important site for exploring gendered dynamics, as platforms present a new public forum within which people attempt to imagine different ways of relating to one another and adds another potential space to perpetuate or contest prevailing gender relations.

As a path into this complex and often impenetrable topic, each of the following examples points to the type of deeper (infra)structural issues that the report explores. Each of these examples highlights how social media infrastructures and gendered experiences intertwine.

- In 2021, Hannah Paranta, a Somalian women's rights activist, was restricted by Facebook from posting content when anti-gender activists conducted a targeted campaign that repeatedly **flagged** her content as inappropriate.
- In 2008, four years after Facebook launched, existing users who had not selected a specific gender identity on their profile were sent a direct request to rectify this, which consisted of two binary options. Six years later (2014), the platform expanded the number of gender options from 2 to 58, providing a much larger list of categories that people could use to self-identify.
- In Uganda, throughout the 2010s, queer communities opted to use dating applications (apps) to manage everyday activities and relations in a context where homosexuality is criminalised, finding this to be a safer space to interact as a community because dating apps fell beyond the radar of government authorities.

The example of Paranta indicates how certain content moderation processes can contribute to the silencing of feminist activists if disingenuous complaints are used as reasons to restrict user access or remove content.

Facebook decided to increase the number of gender categories available to users as a response to feminist and queer critiques. However, this change was misleading, given that in the back-end **dataset** (which is used for content targeting) the options continued to be reduced to only a few. This raises questions about why the options were changed, and whose interests drove these changes – first from non-mandatory to mandatory fields, and then from a few to many gender options for users.

The case of Uganda exemplifies how social media platforms have provided, at times, spaces for groups who are traditionally marginalised or discriminated against to convene together. For queer communities, certain social media networks have allowed users to reduce their exposure to gendered discriminations and navigate contexts where they face criminalisation for their sexuality. This is important from a gender lens as it illustrates how dating apps or private social media channels offer an alternative public sphere that is comparably safer for non-**heteronormative** or gender non-conforming individuals.

As indicated in these examples, the *infrastructure of social media platforms* (see Box 1) is bound up in gender-based experiences. The extent to which platform infrastructure constrains and directs how users are likely to interact online, and with each other, has become an increasingly heated topic of concern. In recent years, increasing evidence has emerged that indicates biases towards false, sensationalist and hateful content on social media. Multiple warnings about bias surfaced in 2021 alone, including former Facebook product manager Frances Haugen's whistleblowing accounts about how the platform's design prioritises profit over people's safety (in this case, specifically young girls' mental health), to Twitter's own research showing that it favours right-leaning political content (Telford, 2021).

### Box 1: Defining the infrastructure of social media

Social media platforms seem to be a growing part of the infrastructure of our daily lives, particularly in terms of how people interact and communicate. Yet, these platforms also have their own infrastructures that shape their design and how they are run (Plantin and Punathambekar, 2019).

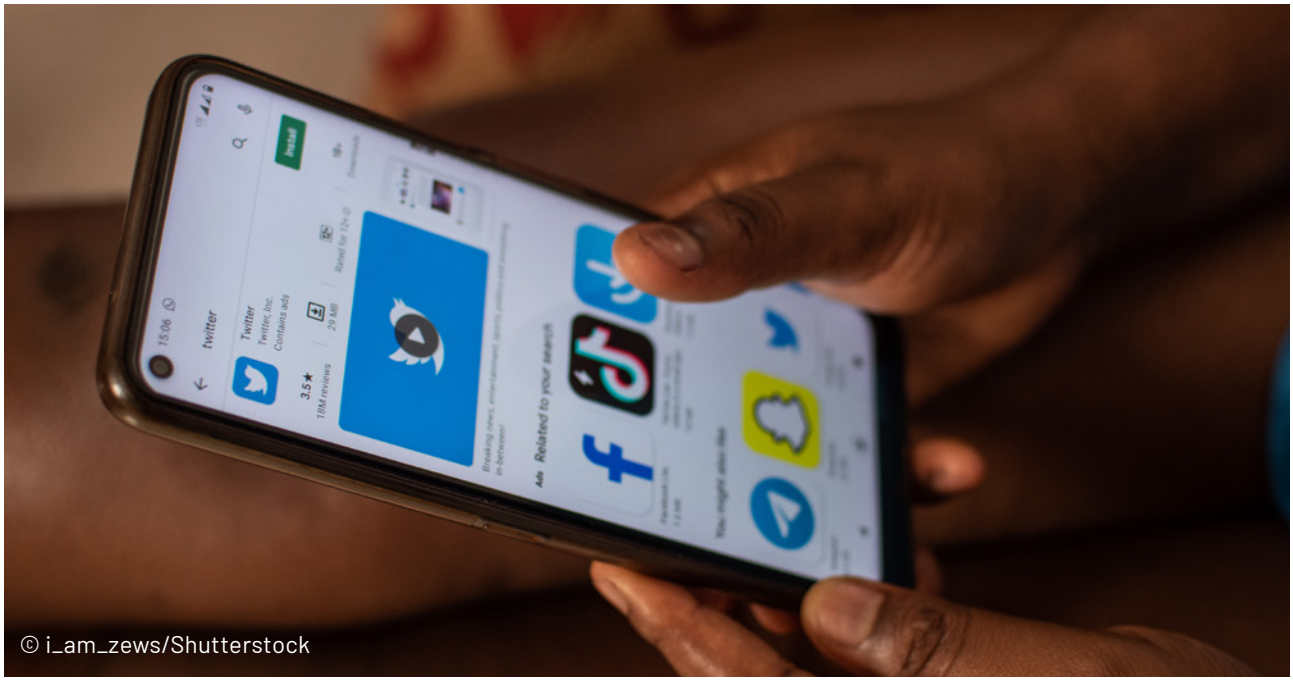
By looking at social media infrastructure, this report considers how *technological*, *organisational* and *political* properties shape the way that social media platforms operate (see also Larkin, 2008), and in turn, what this means for the reproduction and challenging of gender norms through their operation and use.

This report sits within a wider turn in media studies of looking at the infrastructure of communication networks (e.g. Parks and Starosielski, 2015), taking into account how the social, material, cultural and political dimensions of social media platforms underpin and shape communication networks.

While infrastructure can have many dimensions, this report focuses on the core economic, technical and organisational components that shape how social media platforms operate.

This report therefore sets out to explore how the economic, technological and organisational infrastructures of social media platforms are influencing gender norms. It is vital to understand the relationship between social media and gender norms. How people perform and enact gendered social rules is inevitably influenced by the spaces in which they do so – in this case, the virtual 'public squares' of social media platforms. The report also poses questions about the ways the ongoing operation of the platforms – for example, the way that designers and company leaders make decisions (Burgess et al., 2016; see also Duguay, 2016) – might be linked to the gendered ways that people interact or are encouraged to interact on the networks.

This report reviews and synthesises interdisciplinary scholarship on gender norms and social media infrastructure. It develops a framework through which to understand the relationship between gender norms and the infrastructure of social media platforms, as well as identifies opportunities for further research. It aims to make the evidence accessible across disciplines to support wider discussions about the challenges and possibilities for generating more gender-equal relations on these platforms.



© i_am_zews/Shutterstock

## 1.1 Gender norms

Gender norms are informal social rules that express how people of a particular perceived gender are expected to behave. Gender norms are often framed in binary terms (female and male) and erase non-binary or gender-fluid identities (ALIGN, n.d). Of course, gender is only one aspect of people's identities, and gender norms intersect with other norms and inequalities around age, race/ethnicity, class, disability and so on. Attending to the **intersectionality** between sexism and other forms of oppression, such as racism, is critical to understanding how gender norms apply differently across different social groups and identities (Crenshaw, 1991). Gender norms both influence everyday interactions and are embedded in the institutions that affect our lives, including social media. They are often deep-seated and taken for granted, and their influence on attitudes and behaviour is not always apparent. Discriminatory gender norms reflect and reinforce power inequalities, often disadvantaging women, girls and people of diverse, non-conforming gender identities. Discriminatory gender norms are embedded both in everyday interactions and in the fabric of social media platforms, and this influences who is seen and heard online, the different experiences of different groups and the most promising ways to prevent online harm.

## 1.2 Social media platforms

Digging into how social media platforms relate to gender norms is not straightforward. The challenge begins with defining what is a 'social media platform'. The scope of what is considered social media has changed over time (Highfield, 2016). Technological change has given way to a range of digitally mediated forms of communication, including blogs, blogging platforms, social networks, content sharing sites, forums and communities, mobile apps, messaging apps and internet-enabled sites. Definitions of social media platforms can focus on different elements: message construction, specific devices or the types of interactions they enable (Rhee et al., 2021).

For this report, a broad view of social media platforms is taken, working with Carr and Hayes' (2015: 50) definition of social media platforms as:

> ... Internet-based channels that allow users to opportunistically interact and selectively self-present, either in real-time or asynchronously, with both broad and narrow audiences who derive value from user-generated content and the perception of interaction with others.

The emphasis on the perception of interaction is important here, as it allows us to consider platforms with different infrastructural arrangements, and to interrogate how different technical designs, business models and organisational structures facilitate and determine interactions online. This report focuses on some of the more popular platforms, namely Facebook, LinkedIn, TikTok, Instagram, Tumblr and Twitter. It also discusses messaging applications such as WhatsApp, Signal and Telegram in so far as they include public messenger groups and include some scope for linking up to others within the system. While it is one of the most popular platforms in terms of number of users, WeChat is excluded, as much less evidence was found on gender norms for this platform. It would be valuable to conduct a comparative study of the underpinning infrastructure of platforms originating in different countries.

> As people interact through the front end of the platforms, they create data trails.

While the facilitation of social interactions is core to what makes social media platforms distinct, it is only the tip of their activities. Social media platforms have multiple, intersecting interfaces. The front end of the platform is where users communicate with one another and create, access and share content. As people interact through the front end of the platforms, they create **data** trails – in other words, information that is generated and recorded from our digital activity. Data is used to inform the design of the platform itself, from targeted advertisements to **filtering** of content. The back end of the platform is where the interface's design is programmed, and determines what users see.

The activities of social media platforms sit within a wider digital ecosystem. The running and use of the platforms rely on different forms of software and hardware.[1] Temporary outages of services, as occurred with Facebook and its applications for approximately five hours on 4 October 2021, are reminders that social media platforms are part of a wider technical system (Martinho and Strickx, 2021). The internet is a network of networks, with **protocols** and systems (e.g. domain name systems, border gateway protocols) that mark the location of sites within the internet and determine efficient ways of moving between different autonomous systems (such as Facebook and Amazon). In basic terms, in the October 2021 outage, Facebook's servers could no longer be found by other networks on the internet, and therefore Facebook was inaccessible to users. Fixing the problem appears to have required a technical team to manually reset Facebook's servers in-person in California (Taylor, 2021).

Social media platforms are used differently by different sectors of society. For instance, platforms can be used by start-up companies for business and marketing, or by governments for public health campaigns. These different uses share a reliance on data analysis and sharing, which tailor platforms to areas of activity. Social media platform data is captured, sorted and processed to direct users' activity and analyse the effects. Van Dijck (2020) argues that social media platforms are at the core of a highly interdependent digital system, a feature that will be explored in this report.

In summary, social media platforms can be examined from different perspectives. Platforms are guided by specific business models, usually designed to make profit. They are software programmes or applications that are continually being updated and linked to other websites and applications. They are spaces where people communicate for diverse social, business and political reasons. Finally, they are organisations with policies, workforces and hierarchies.

1 For example. including coded protocols, internet service providers, app stores and pay systems such as PayPal and Apple Pay.

Each perspective offers different insights into how the infrastructures of social media platforms relate to gender norms:

- as technologies based on decisions made by software engineers, designed to engage with people's data in particular ways (e.g. through the embedding of certain gender categories or biases into how the technology runs)
- as organisations, in which organisational cultures, leadership and workforces shape decision-making about the technology (e.g. terms of use, company strategies and values)
- as spaces for people to communicate and engage with information (e.g. through responding to digital content from other users)
- as agglomerations of different interests – for example, those of users, shareholders, Chief Executive Officers (CEOs) and advertisers – informing choices about what content is visible, amplified and removed
- as entities that are subject to external regulations and policies, which can be used to encourage or institute changes in gender norms.

## 1.3 Methodology and report structure

This report unpacks how the infrastructure of social media – broadly defined as the economic, technical and organisational structures that underpin a platform – shapes the expression and contestation of gender norms online. It complements a forthcoming ALIGN report that delves into the potential of online activism to shift gender norms (Washington and Marcus, 2022) by exploring how both feminists and anti-gender activists use social media platforms to further their cause, as well as whether the platforms themselves contribute to or constrain the power to instigate gender norm change.

It is based on a targeted review of academic and grey literature on gender identity, sexuality and the infrastructure of social media, identified through a combination of systematic search queries and snowballing. Given this, the report tends to focus on Meta (formerly Facebook) and its core products (Instagram, Facebook, WhatsApp and Messenger), due to its dominance globally (reporting 3.45 billion people using at least one of its core products each month in early 2021) (Meta, 2021a). It is made clear throughout when findings are specific to Meta products.

The report aims to present a framework for exploring the relationship between gender norms and social media infrastructure that is accessible across disciplines. It is a starting point for further discussion and research on the relationship between gender norms and social media platforms from an interdisciplinary lens that looks at not only what users do, but also how the infrastructure is designed and how it works.

The report is structured as follows: first it explores how social media infrastructure works from economic, technical and organisational perspectives. It then unpacks the evidence for how this infrastructure relates to gender. The fourth chapter turns to the regulation of social media, and how this has intended and unintended consequences for gender norms. The final chapter concludes with recommendations for further research.

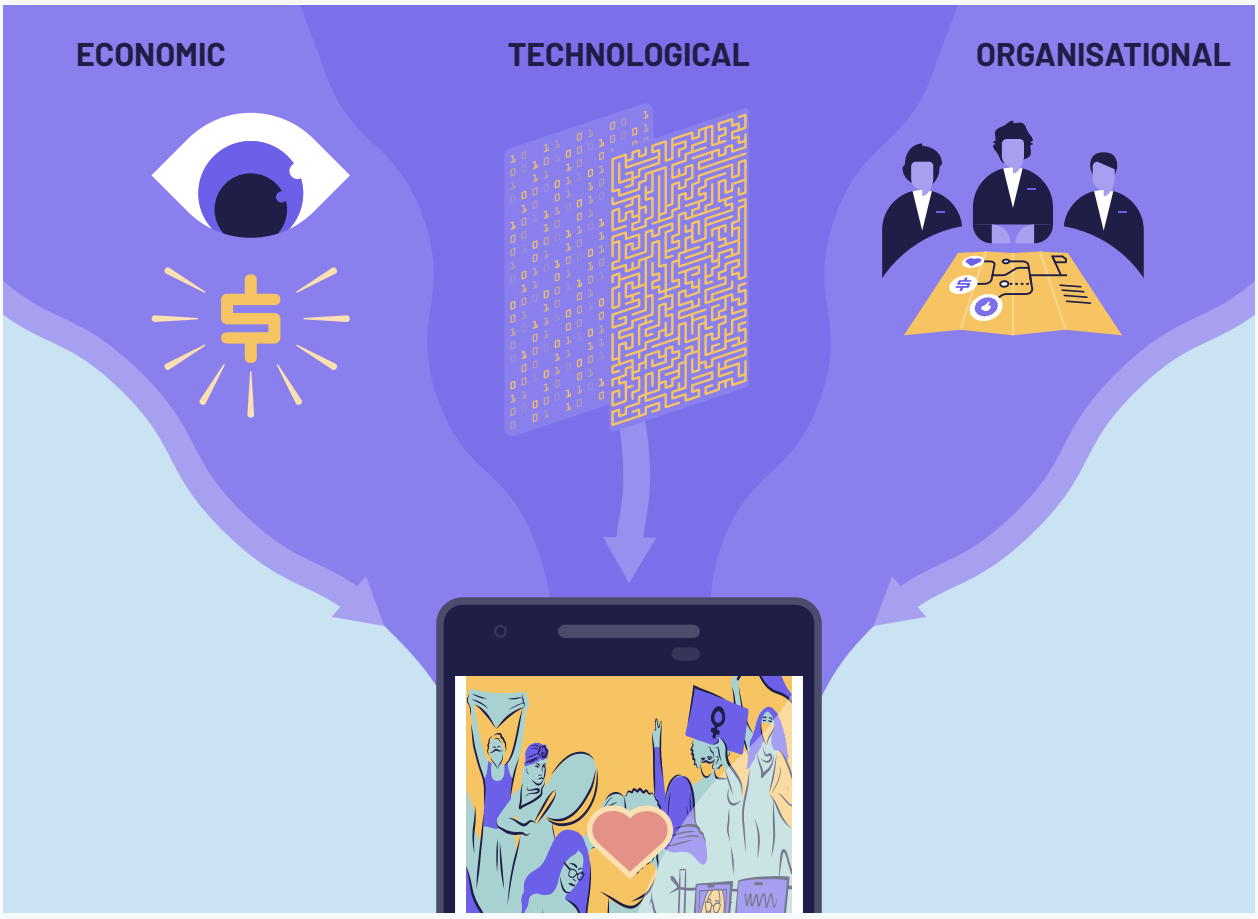
# 2 Social media infrastructure: how does it work?

How do social media platforms work? The infrastructure that shapes how social media platforms operate has different layers (see Figure 1), including:

1. economic: how platforms are sustained economically, usually for profit
2. technological/material: the software that makes up the platform
3. organisational: the structure and culture of the company that makes and runs the technology.

Other factors also influence how social media platforms operate, such as the physical hardware (e.g. servers) required for software to run. The report focuses on the three dimensions listed because they directly relate to the types of interaction and content that appear on the platforms, and therefore appear to be most relevant to the construction of gender norms. This chapter introduces each of these components, providing the basis from which to investigate the relationships between platform infrastructure and gender.

Figure 1: Hidden layers of social media infrastructure



## 2.1 Economic infrastructure: social media platforms' business models

The early 2000s saw the rise of some of the social media platforms that dominate today, including LinkedIn (2002), MySpace (2003), Facebook (2004), Reddit (2005) and Twitter (2006). Their transition into profit-generating enterprises took place over the next few years, as platforms took up different opportunities to monetise user engagement. They did this both directly and indirectly through the data produced from user activity – for example, through the introduction of advertisements on platforms.

Since the early 2000s, several social media platforms have generated high levels of profit. In 2020, Facebook's net income rose 101% to $10.3 billion, with a net worth of over $1 trillion (Murphy, 2021). Twitter's net worth in mid-2021 was estimated at $4.4 billion. By contrast, other platforms have not been so profitable. Tumblr, for example, fell in value; it was bought by Yahoo for $1.1 billion in 2013 but was sold six years later for only $3 million. Tumblr still aims to generate greater profit; in July 2021 it introduced a beta version that gave some users the option to monetise their content through a paywall accessible only to subscribers (Siegel, 2019). The difference in Tumblr's business model and technical structures, which has given way to different possibilities for people's interactions, is explored throughout this report (see in particular Sub-section 3.3).

### Profit generation models

Although some platforms are much more effective than others at generating profit, most platforms tend to look to user data and engagement activity as the basis of profit generation.

Some scholars argue that social media platforms commodify people's social interactions. Dean (2008) situates social media companies within a wider trend that turns *communication* into something that can be sold: more clicks, more connections and more sharing bring more value to those who want their information to get the attention of a particular/growing audience. In *The age of surveillance capitalism*, Zuboff (2019) suggests that social media companies take part in the commodification of data – data that is produced as a result of our online presence and activities. Companies process data and use it to target products and marketing campaigns. Third-party advertisers pay to access advertising space aimed at target audiences. Zuboff argues that social media companies are data companies, primarily serving advertisers by providing them with opportunities for targeted advertising (see also Alaimo and Kallinikos, 2017).

Opinions differ about the extent to which this model is exceptional, or even unfairly extractive. Some contest the extent to which an *individual's* data contributes to value creation compared with, for example, the value generated from aggregated data (e.g. Gilbert, 2021). Couldry and Mejias (2019), conversely, discuss the use of people's data for profit as a wider problem, using the concept of 'data colonialism' to highlight how people are exploited through the use of data for profit. Similarly, Morozov (2019) also sees surveillance capitalism as an extension of contemporary capitalism, tied to longer, systemic economic forces rather than being an exceptional form tied to the choices made by tech companies.

Irrespective of the level of concern or perceived exceptionalism of a data-driven profit model, the use of data from user profiles and behaviours online to assist advertisers is core to social media companies' profit. The scale of advertising revenue supports the view of social media platforms as data companies. In 2020,

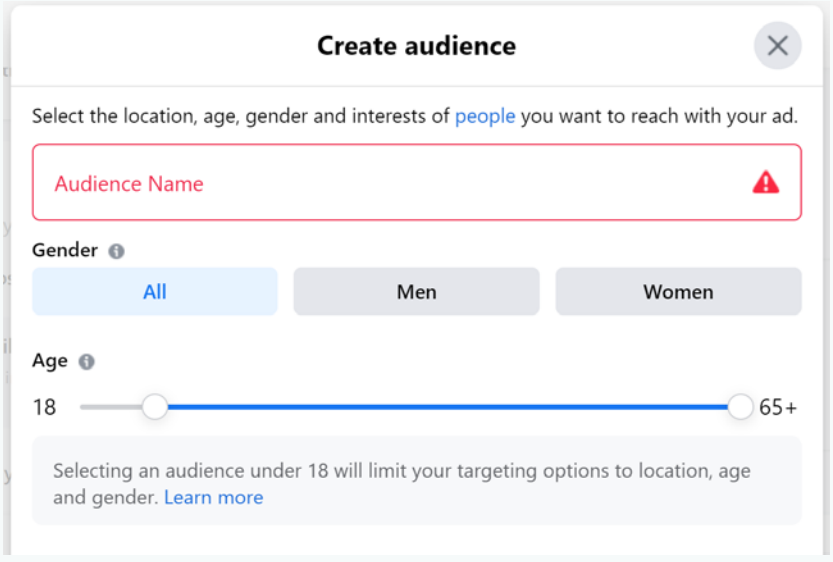
Facebook and Twitter generated $84.2 billion and $3.2 billion in advertising revenue respectively, which represented 97.9% and 86.3% of their total revenue (Van der Vlist and Helmond, 2021).

## Creating and monetising data

Data is key to a profit model based on targeted content, as effective targeting assumes the accurate, or at least useful, representation of people as data. Information about users' sex and gender has been translated into data as part of these processes.[2] A distinction is required between how a user self-presents on the front end of the platform and how they are categorised in data. On the front end, users have faced different options for identifying gender. Platforms such as LinkedIn and Twitter do not require users to identify their gender, and users can choose not to select a pre-defined gender category. Facebook, by contrast, has required new users to identify their gender and select preferred pronouns since 2008. Existing users who had not previously identified a gender received a prompt in 2008 requesting them to select a binary option.[3] These users could choose to continue not to identify their gender. However, users who had previously identified as male or female, along with new users, were required to select a binary option in order to use the platform. As mentioned in the introduction to this report, the options available to users have since expanded.

Flexibility in how users identify on the front end of the platform contrasts with what exists on the data-side of the platforms, where gender is still grouped into more general categories. Facebook has retained only three options for advertisers and third-party applications: female, male and undefined (Bivens, 2017: 886). For example, the basic 'create ad' function on Facebook offers 'All / Men / Women' as the standard options for audience targeting (see Figure 2).

Figure 2: Screenshot of audience options for Facebook's 'Create Ad'




*Author's account, 14 February 2022*

2 Users face distinct limitations in categorising their gender on relationship and dating applications. For example, when creating an account on okcupid.com, users are given choices between a binary gender and one of three orientations: 'straight', 'gay' or 'bisexual'. On eharmony.com, users can only identify as a man or a woman, and can only express interest in either men or women (McNicol, 2013).

3 Bivens (2017) suggests that Facebook made gender a mandatory field as part of its monetisation strategies, i.e. because the information was useful to third parties that used Facebook's user database.

Platforms such as Twitter, which does not require users to identify their gender, infer gender from users' activities and use this to profile/categorise users. The 'Twitter for business' page has stated, 'Gender is determined via public signals that users share on Twitter, such as @usernames or accounts followed' (2015, in Bivens and Haimson, 2016). A classification system (the technicalities of this are discussed in Sub-section 2.2) is used to compare user data to a feature dataset (see also McNicol, 2013).

Over time, in order to prevent discrimination, social media companies have had to place some limits on the types of data that can be used to inform targeted advertising. Previously, Facebook's advertising could also be targeted by sexuality, gender and race, among other characteristics. Race and ethnicity were classified under the category 'multicultural affinity', which included sub-categories such as 'Hispanic (US – All)' or 'Hispanic (US – bilingual)'. White or Caucasian was not an explicit option, but could be targeted by excluding all other categories, indicating how whiteness is taken as the default setting or as the absence of an additional identity (Sances, 2021). Racial and ethnic targeting options were removed after exposés showed targeting could discriminate against people of colour in housing, jobs and credit advertisements (Hao, 2021).

Within this data-based model for profit generation, sex remains one of the **data points** used to target content. This indicates a direct way that particular constructions of gender might be 'baked in' to the infrastructure of the platform, as different types of content are targeted to users based on their assumed sex (Bivens and Haimson, 2016). Evidence on how this has contributed to user experiences is explored in Chapter 3.

## Business partnerships

Partnerships between social media platforms and third-party applications and websites expand opportunities for profit generation. Partnerships reduce friction in data sharing and processing for businesses and platforms. For users, this means a front-end experience in which different websites and applications can be more individually tailored and linked up. Partnerships expand the services that platforms can offer to businesses, such as chatbots, payment integration and automated messaging. Platforms can also provide and combine more data on customers, collecting and storing data on audiences and prices (Van der Vlist and Helmond, 2021).[4] These types of combined dataset provide even more granular insights into a person's preferences, and enable the partners to target users on multiple sites. This raises a number of questions around gender. To what extent might data sharing and integration result in more consistency in targeted content across sites, thereby reinforcing specific messages? What does this mean for an individual's experience? Does greater personalisation result in the diversification or homogenisation of gender norms?

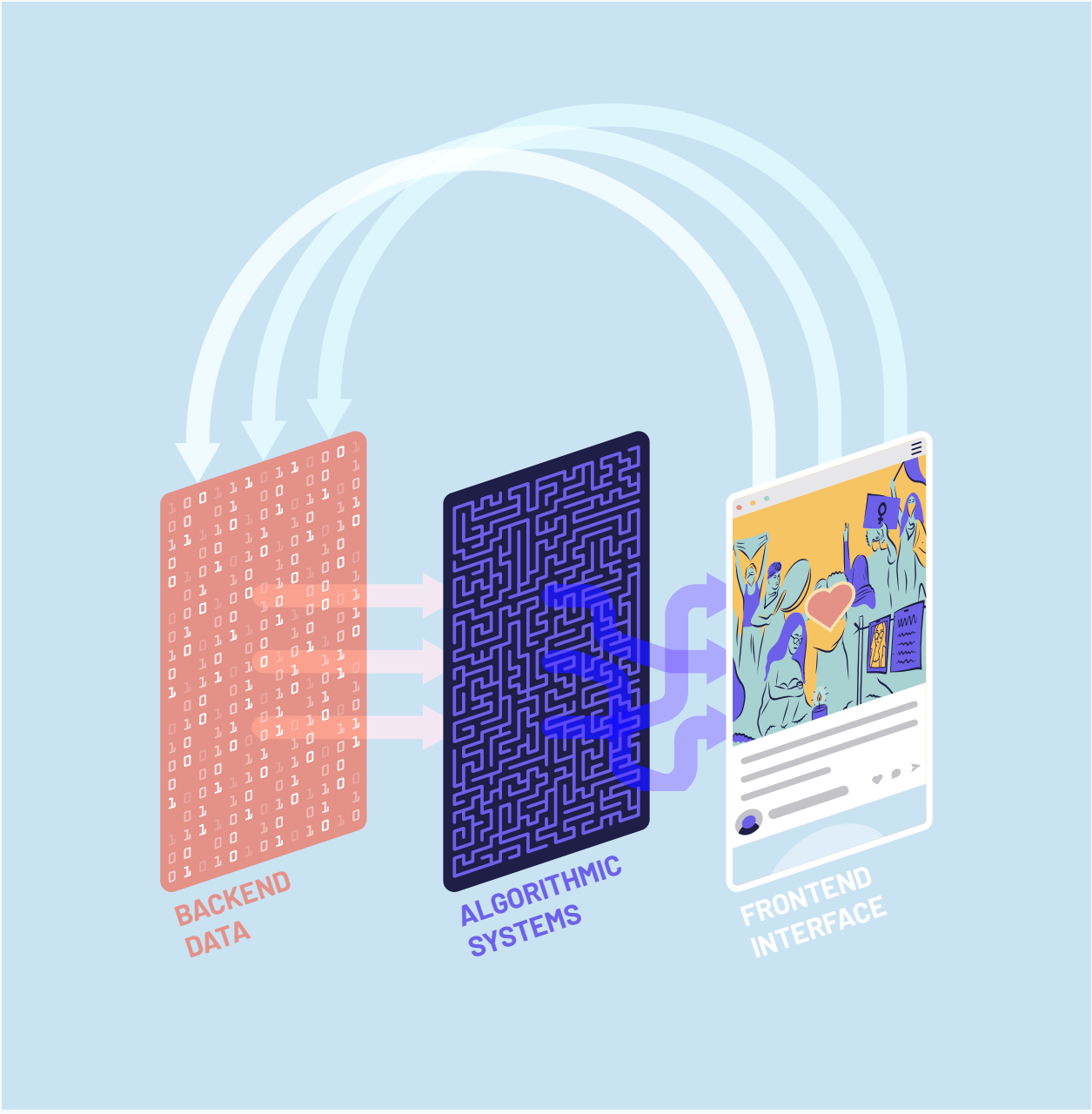
" Racial and ethnic targeting options were removed after exposés showed targeting could discriminate against people of colour.

4 Social media platforms of varying sizes hold different kinds of data. Larger social media platforms, such as Google and Facebook, can access first-party data through their login services. Smaller platforms can use data intermediaries to gain access to second- and third-party data sources (the latter is usually less valuable, as it comes from less known sources).

## 2.2 Software and technological infrastructure

As indicated in Section 2.1, the business models of social media platforms are tied to their technical software. This section provides a series of mini explainers about technical aspects of social media platforms (which form another layer of social media infrastructure; see Figure 3) that are important to both their profitability and to user experiences. The section discusses how gender and sex are accounted for in technological design and other development processes. This discussion provides necessary context for Chapter 3, which explores how the technological features of social media infrastructure shapes people's online experiences and interactions in relation to gender norms.

Figure 3: Technical layers of social media infrastructure



## Digital identifiers

Before they can interact on social media platforms, users must be identifiable as discrete individuals. Platforms have different requirements for individual identification. Most platforms require a person to provide at least an email address or phone number to create a profile, although a few platforms, such as the 4chan message board site, do not require any user information for someone to participate. Some platforms, such as Tumblr or Twitter, require little personal information about an individual for them to create a profile. Fewer information requirements make it easier for a person, organisation or other entity to create one or multiple user identities, and to detach who they are online from who they are offline. By contrast, Facebook is designed to map onto someone's physical person and offline relationships. It is based on the assumption that one Facebook user equals one person. The platform encourages users to connect with people they know from their other social circles, such as workplaces, neighbourhoods, schools and families, thereby potentially blurring online/offline divides (Marwick and boyd, 2011; Dragiewicz, et al., 2018).

## Algorithms and algorithmic systems

Social media companies design software to shape what people see on the platform and how they can use it. **Algorithms** play a critical role in automating decisions about what content to show which user, and in what format. Social media platforms such as Facebook, Twitter and Instagram rely on algorithms to continually curate an individually tailored online environment that presents, amplifies and monitors content while connecting users.

Algorithms are sets of rules and instructions on how to deal with information. Based on these rules, algorithms can respond to data received from their environments, which in turn can enable them to incorporate additional data into their processes and outputs (Lomborg and Kapsch, 2020). Algorithms therefore help to organise and manage platform activity, for example, by helping to determine the order in which someone sees information, or by deciding what content is flagged as inappropriate.

Algorithms are also a critical part of a data-driven profit model, which is based on algorithmic processing of data. Algorithmic systems are programmed to learn from user data in order to make recommendations for targeted content that is relevant to users. This is essentially a technical shortcut for sorting each individual's posts online based on the popularity of posts and their relevance to users. Algorithms are often proprietary (owned by the company), meaning that they can be kept secret. As a result, the functioning of platform algorithms is often inaccessible and incomprehensible to users.

While humans build algorithms, they can be designed to operate independently of human supervision to different degrees. For example, as discussed further in Chapter 3, the identification and removal of inappropriate content can be fully automated or have built-in functions to alert that human review is necessary.

## Digital data in algorithmic systems

With the exponential growth in social media users and activity, automated content moderation is increasingly central to managing what content users see. With a higher volume and more diversity of content, users see a decreasing proportion of content and it becomes more difficult for them to capture other users' attention (O'Meara, 2019).[5] Algorithms must be able to read content in order to process it; in other words, content needs to be turned into **data**. This process makes complex information about the user interpretable and actionable.

Cheney-Lippold (2011: 165) argues that a system based on information as data constrains how gender identity can be conceptualised:

> Online a category like gender is not determined by one's genitalia or even physical appearance. Nor is it entirely self-selected. Rather, categories of identity are being inferred upon individuals based on their web use. Code and algorithm are the engines behind such inference...

Algorithms organise and process data that is derived from the messiness and complexity of human life (Apprich et al., 2018; West, 2020). The nature of the information affects how easy it is to translate that information into data that is readable by algorithms. DeVito (2017) explores how algorithmic systems turn subjective indicators into objective categories. Subjective identities such as gender invoke immediate challenges, as they do not correspond to discrete categories.

Taking the example of friendship, DeVito shows how Facebook has translated this subjective construct into an objective definition. According to Facebook's algorithm, a friend relationship is defined by a connection between two different Facebook users. The algorithmic moderation of Facebook's news feeds is informed by these 'friend' relationships, which then feeds into the algorithmic systems that determine (in combination with other values, such as user interests, engagements, preferences and content quality) what a user sees on their feed. Assumptions made in defining constructs such as friendship can have wider implications in terms of who people are likely to connect with online. Kurgan et al. (2019) show how Facebook algorithms reflect the principle of 'homophily' in friend relationships, where similar people are assumed to like each other. This means that users are more likely to be recommended content that is similar to what they already like. They are also more likely to be prompted to connect with other, similar users, rather than across differences.

These challenges have implications for sex and gender identities. Despite the prevailing dominant understanding of sex and gender as binary categories (male/female and man/woman), there is a well-documented diversity of sex and gender experiences beyond these binaries. The reduction of these complex experiences into a limited set of discrete categories marginalises people whose sex/gender identities do not align with dominant cultural norms. Queer, trans* and non-binary people experience high levels of marginalisation and violence within society due to their perceived transgression of gender norms.[6] When platforms reinforce limited ideas of gender, they erase diversity and contribute to this marginalisation.

5 In April 2010, Facebook identified three variables - objects (e.g. photos), users and interactions (also referred to as 'edges') - that are algorithmically ranked. In this ranking system, the weight assigned to these variables determined the visibility of content (Bucher, 2012).

6 The asterisk in trans* is used to acknowledge the diversity of trans identities and experiences.

Algorithmic processes that prioritise similarity also reduce the likelihood of people interacting with others who identify in different ways.

Still, some uncertainty remains about the extent to which data on gender or sex inform what a user sees. The algorithmic systems that determine how content is filtered and presented to users are increasingly complex, processing, combining and prioritising diverse data points. **Principal component analysis** is sometimes used to isolate the strongest correlations in data. This gives more attention to 'blunt' content (e.g. click throughs) rather than more subtle or complex data about a person's identity. This makes it harder for systems to account for how gender is produced online (Sumpter, 2018, in Gerrard and Thornham, 2020).

Platforms also use processes relying on data and algorithms to identify, review and remove problematic content. Again, this requires content to be transformed into digital data so that it is readable to algorithms. Gorwa et al. (2020) highlight two technical processes used to identify problematic content: hash matching and classification. Hash matching starts with representing content as a string of data (a hash). A hash is a string of data that identifies the distinctive features of a piece of content. This turns the content into a simpler, computationally readable form and makes it easier to single out particular pieces of content, by looking for matches in the hashes. A new piece of content, in the form of a hash, is compared to a database of curated examples. For example, following a shooting at a mosque in Christchurch, New Zealand in 2019, which was live-streamed on Facebook, Facebook created and released hashes corresponding to 800 versions of the shooter's video. The platform used this to quickly compare any new content, singling out any that matched distinctive features of the shooter's video, encased in these 800 hashes (Gorwa et al., 2020). Classification processes use algorithms to identify statistical patterns in data. They make generalisations about the features of new content by making assessments about new content based on models trained from other big datasets. Hash-matching requires a previous version of content, such as the shooter's video. This is not the case with classification. Here, machine learning is used to train algorithms to single out content that is likely to fit a particular category (e.g. violent attacks).

## Application programming interfaces

Partnerships between social media platforms and other applications (discussed in Sub-section 2.1) require technical integration. Technical interfaces and packages facilitate integration and sharing between applications and sites. **Application programming interfaces** (APIs) are interfaces between, for example, the social media platform and other third-party applications. APIs allow the platform and external applications to speak to one another and share data.[7] APIs can set different levels and types of access for different actors. Bodle (2011: 333) explains how APIs can operate as a form of 'controlled openness' for the platform:

> Facebook's development and integration of Open APIs demonstrate a strategy of controlled openness, innovation, and dominance. Facebook utilizes Open APIs to provide valuable user data to developers and online partners in order to encourage the proliferation of social applications that ultimately harness inbound links from external sites and devices by redirecting traffic to Facebook's servers.

7 **Software development kits** are more comprehensive packages that support the integration and communication of different applications (Van der Vlist and Helmond, 2021).

A social media platform's size and influence with regard to its partners affect the levels of control it has over the terms of the partnership. This can work both ways; other organisations' standards and processes can influence how the platform operates. In November 2018, Apple banned Tumblr from its **app store** due to concerns about pornographic content. This was soon followed with a decision by Tumblr to ban sexual content. Some argue that the regulation of data sharing is likely to have a greater impact on smaller platforms that rely on external parties' **data sources**, with less effect on large platforms, such as Facebook, that produce more data internally (Van der Vlist and Helmond, 2021: 12). From a gender norms perspective, this means that larger and more influential partners' approaches to sex and gender identities could matter beyond the platform itself, depending on how this affects the terms and operation of partnerships.

## 2.3 Organisational structure

Organisational structure forms a third layer of social media infrastructure (see Figure 1). Behind the software, social media platforms are companies. To understand what is possible on the platform, the organisational structure, hierarchy and culture needs to be taken into account. Company stakeholders, from CEOs to engineers, make decisions about organisational and technical priorities that have implications for how people might interact through the platform.[8] This section focuses on existing evidence about how gender and sexual identity are reflected in the organisational dimension of social media platforms. This dimension forms a third set of considerations, alongside profit models and technical structures, that frame the possibilities for what takes place on the platforms. Attention to the organisation itself is critical for understanding how and why social media platforms develop in specific ways, and why certain opportunities, processes and types of resistance emerge.

Most large companies face challenges in fostering diverse and equitable workforces, and concerns have been raised more specifically about the lack of diversity in the tech sector. This remains true for social media companies. While there is little research on social media companies' workforces, company data indicates a lack of diversity, specifically in leadership and engineering roles. Men have the most influence over platform design, while women and marginalised groups and individuals are more often assumed to be end users: targeted by advertising, or creators of user-generated content (Gregg and Andrijasevic, 2019). Companies recognise diversity is a challenge, and most have responded by publicising their equality, diversity and inclusion initiatives. Yet, with few exceptions, men continue to dominate in leadership and technical roles (see Sub-section 2.3 for more details on race/ethnicity).

### Decision-making power

The priorities of social media companies are shaped by how decisions are made and who has influence. These priorities also frame decisions about content moderation and user experiences. From a gender perspective, therefore, who has influence, and the biases that emerge in decision-making processes, matters.

8 Eubanks (2018) and Noble (2018) have drawn attention to implicit biases in the outputs of algorithmic moderation more widely.

The power of some CEOs and founders of social media companies in company decision-making highlights the unique individual influence over some platforms' priorities. Mark Zuckerberg, founder and CEO of Facebook, is at one extreme.[9] Zuckerberg's outsized influence provides an example of a founder-led organisational culture (Moran, 2020). The organisation overall is divided hierarchically and divisionally into function-based, geographical and product divisions. This provides a structure with a top-down decision-making hierarchy. Zuckerberg is chair of Facebook's eight-person board of directors, CEO and also maintains majority voting power.

"

While men do dominate social media companies, this does not mean that women are absent from their decision-making bodies and leadership positions.

Still, his influence is not without constraints. Financial viability and profit matter to Facebook and, to some degree, Zuckerberg must contend with other shareholders. For example, in 2016, Zuckerberg attempted to introduce a third share class, on top of its dual-class shares in which shareholders had different votes per share: one class with one vote per share and another with 10 votes per share (Stone, 2009). This new class would be a non-voting share class and would allow shareholders to sell shares while retaining voting power. The initiative faced minority shareholder resistance and was withdrawn (although resistance was not cited as the reason for the withdrawal) (Constine, 2017).

While men do dominate social media companies, this does not mean that women are absent from their decision-making bodies and leadership positions. While independent of the company's internal decision-making process, Facebook's Oversight Board (see Box 3) includes men and women from diverse professional, geographical and cultural backgrounds. Also, social media companies have been pushed by legislation in the national/state jurisdictions where they are based to improve the representation of women. In 2018, the state of California mandated that state-based public companies would be required to have at least one woman on their board by the end of 2019. This generated a response among tech firms in and around Silicon Valley, including social media companies. Overall, 126 tech companies recruited 138 women to their boards (Sonnemaker, 2020). Around this time, in March 2020, Facebook announced two new female members to its board of directors, making it 40% women (4/10) (Bursztynsky, 2020).

Decision-making power is not simply a question about numbers. Given how organisational priorities have shaped the way that platforms operate (as will be discussed in Chapter 3), it is clear that decision-making processes have an impact on and are impacted by gender norms. Recent efforts by companies indicate some effort to at least alter the gender balance within leadership roles. However, evidence suggests that platform priorities can impact gender norms and even harm certain users. This indicates a need for more research into how decision-making structures might shape company priorities and platform design.

9 Not all founders and CEOs of social media platforms retain as much control as Zuckerberg. For example, in early 2022 Twitter co-founder and CEO Jack Dorsey only holds 2.34% shares in the company (Walton, 2022).

## The composition of the workforce

Gender discrepancies are visible throughout social media companies, particularly in leadership and technical roles. Despite companies' equality and diversity initiatives, women remain a minority and there are gaps in pay and bonuses. Table 1 shows relative consistency in the proportion of women overall, and in technical and leadership roles, across a sample of social media companies.

**Table 1: Proportion of women in the work force: social media platforms**

| | Overall | Technical roles | Leadership roles |
|---|---|---|---|
| Twitter (June 2021) | 43.7% female/women | 29.2% female/women | 37.7% female/women |
| Facebook (2021) | 36.7% female/women | 24.8% female/women | 35.5% female/women |
| LinkedIn (2020) | 44.7% female/women | 24.0% female/women | 41.9% female/women |
| TikTok | No data found | No data found | No data found |

*Sources: Twitter (n.d.); LinkedIn (2020); Williams (2021).*

*Note: As the organisational structures for leadership and technical roles differ in each company, these numbers are not directly comparable.*

Taking Facebook as an example, the number of female employees has risen in the past year, but only from 36.9% to 37.0%. The proportion of women in technical roles was even lower, increasing from 23.0% to 24.1% (Statista, 2021a).

Social media companies are not representative of wider populations in relation to race and ethnicity either. In 2021, only 4.4% of United States (US) Facebook employees identified as Black, compared with 45.7% as Asian and 39.1% as white. The type of roles also varies by race/ethnicity, with 60.9% of leadership roles held by white employees, and 54.4% of technical roles held by Asian employees (Facebook, 2021). These statistics provide only a partial window into bigger questions of representation across social media platforms' workforces. While more granular statistics across specific roles (e.g. content management, hardware versus software engineering) are mostly unavailable, indications suggest there are demographic variations in terms of who works where. This would be a valuable avenue for researchers to explore, in order to make sense of how and why different biases in platform operation might emerge.

There are also disparities in pay. Social media companies are not necessarily exceptional in the existence of gender pay gaps. For example, in April 2020, Facebook reported a mean gender pay gap of 5.1% and a median gender pay gap of 11.4% in the United Kingdom (UK), compared with a national median of 14.9% (for both full- and part-time workers) (ONS, 2021). However, bonuses must also be taken into account. In the same year, women's average bonuses from Facebook were 41.8% lower than men's. In another example, TikTok, a social media platform based on music and short videos, reported a mean gender pay gap of 33.0% and a median pay gap of 14.1% in April 2020. Again this discrepancy extended to bonuses, with average bonuses for women being 36.1% less than for men (a median gap of 73.1%) (TikTok, 2021).

Facebook suggests that differences in pay and bonuses are partially attributable to the types of job that women occupy, specifically their lower representation in technical and senior technical roles (Meta, 2020). There are exceptions, see Box 2. Still, the point here from a gender norms perspective is not whether the difference is exceptional, but if and how the nature and conditions of the workforce shape how the platforms operate, and what biases might go unseen. This question has not yet been directly addressed in research on gender and social media platforms. Such research would need to investigate the internal organisational practices that give way to specific decisions about platform operations.

### Box 2: Bumble: does this female-led dating app challenge gender norms?

The dating app, Bumble, was founded by a woman and markets itself as a company that is kind to its workforce. It has an 85% female workforce, and a fund that invests mainly in start-ups led by women and underrepresented groups. Although little research has been done so far, Bumble provides an opportunity to explore the relationship between a female-led workforce and user experiences on the platform. For example, while Bumble claims to give women agency on its app, paralleling its gender-friendly ethos, the effects of this are mixed and have generated some criticism. The app requires women to message men first, and some suggest that this compounds inequality in labour, and that it does not necessarily challenge dominant gender dynamics and ideas of attractiveness.

*Sources: Bumble (n.d.); Strimpel (2021).*

## Contract and task-based work

Contract and **crowd workers** play a key role behind the scenes of social media companies, particularly in the creation of labelled datasets and the training of algorithmic systems. These, in turn, shape which content appears to which users and which content is amplified, demoted or removed.

Social media companies often work with contract and crowd workers on specific tasks, such as data labelling. This work sits within a wider 'gig' economy, through which individuals obtain work on a per-task basis, often working remotely. While this expands work opportunities to diverse locations globally, workers within this digital gig economy have little opportunity to organise, and often work under tight time constraints with low earnings (e.g. Posada, 2021).

Outsourced work is critical to the functioning of the **machine learning** systems that underpin social media platforms. This work often includes the task of labelling data to train machine learning algorithms. The use of training sets to teach machine learning algorithms how to read content is ubiquitous in the world of machine learning. Algorithms need to learn to recognise and differentiate data in order to process and filter content. At the most basic level, this involves a person describing a piece of information, and then this being done repeatedly for many different labels and huge amounts of content. Through this, algorithms learn to distinguish and associate content with different classifications. When labelling data, crowd workers have to make judgements about the world, people's environments and their appearances. This removes users' power to self-identify; instead, they are labelled by an unseen, often poorly paid and widely distributed workforce (Crawford and Paglen, 2019).

The individual judgements of data labellers at the heart of the task of labelling training datasets is hidden by the scale and size of this activity. For example, Google has an Open Images Dataset with 9 million images and a video database with YouTube that has 8 million labelled videos. These datasets were created by nearly 50,000 workers who sorted and labelled pictures (Reese and Heath, 2016).

Little is known about the make-up of contracted and task-based work, including what attention is given to equality, diversity and inclusion. Yet, these workers are required to bring their own judgements to classifying information, including about people's appearances, identities, emotions and attitudes. As Chapter 3 discusses, the output of algorithms trained through these labelled training datasets has led to gender-related biases. This suggests that more attention is needed to the minute, subjective decisions that are being made about people and how different judgements come to bear.

## 2.4 Reflections

This chapter has provided an overarching picture of some of the key layers of social media infrastructure, exploring how and why platforms operate in particular ways. The chapter began with the business model through which platforms become profitable, before turning to the technical structures that underpin profit generation and user experiences. It concluded by reflecting on social media platforms as companies with workforces and organisational cultures.

Each of these (economic, technological and organisational) layers of social media platforms impact on and are impacted by gender norms. They also interact with and impact each other in deep and complex ways. The dominance of men in both leadership and technical roles undoubtedly affects both software outcomes and business decisions. The fact that platforms (including algorithms for data labelling and content moderation) are predominantly designed by men, with women and marginalised groups presumed only to be end users or content creators, raises many concerning questions in relation to gender and other forms of equality.

These questions require further research. For instance, research on crowd work could explore whether changes in who designs, tests and tweaks algorithms, and under what conditions, could help to identify gender biases and adjust them early on in their design and operation. Growing attention to gender imbalances indicates the potential for change, evident in increases in the number of women employees, however gradual. Yet, gender inequality is much more than a question of numbers. Better representation of women in both technical design and positions of leadership is needed, but so too, crucially, is a deeper understanding of the gender (and broader equality) politics of data. Without this, platforms will continue to reinforce dominant and patriarchal gender norms.

> The dominance of men in both leadership and technical roles undoubtedly effects both software outcomes and business decisions.

# 3 Social media platforms and their influence on gender norms

As discussed in Chapter 2, the infrastructure of social media platforms consists of different (economic, technological and organisational) layers. Each layer provides a different entry point for understanding the gender dynamics of social media. This chapter examines how the features of these different infrastructural layers shape people's experiences on the platforms. It aims to answer: *How does social media infrastructure contribute to expectations about how people should behave?* To answer this question, the chapter explores existing research into how social media infrastructure – business models, organisational structures and technology – relates to user experiences and shapes gender norms. It also considers the implications for gender equality and activism, and the scope for challenging or changing gender norms through social media interactions.

The chapter discusses four areas of research: (1) algorithmic outputs driven by profit incentives that promote images of gender that are viewed as profitable; (2) algorithmic processes themselves, experienced by users as a form of disciplining power on preferences and identities; (3) the **influencer** industry, and user involvement in profit-making through attention online; and (4) social media companies' community standards, and how companies identify and enforce problematic activity.

## 3.1 Gender norms and the outcomes of technical design

### Automated outcomes

The previous chapter explained the basic ways that algorithms are used in social media platforms to prioritise content and moderate what users see. It showed how the design and running of algorithmic systems is not neutral: it involves judgements about how to label content to train algorithms and what rules should guide how algorithms moderate content. Given this, it is perhaps not surprising that studies of the outputs of algorithmic processes on social media platforms show that they are biased along gender lines. Users have been found to disproportionately see content on social media platforms that reflects prevailing patriarchal gender norms.

Noble's (2018) seminal work on **search algorithms** brought attention to gender bias in algorithmic outputs. Noble exposed how Google's search algorithms retrieved hypersexualised and pornographic content for searches for women of colour.[10] Similar patterns have been found on social media platforms (Gieseking, 2017; Bishop, 2018; Roberts, 2018). In 2021, biases in algorithmic content moderation on Instagram generated public controversy, when Frances Haugen, mentioned in the introduction to this report, handed over internal Facebook documents. These included studies by Facebook that found that 13.5% of UK teenage girls felt that Instagram worsened their suicidal thoughts and 17.0% of teenage girls said their

10 Over time, algorithms were corrected to exclude this content, first for Black women (specifically for the search term 'black girls') and then later for other groups, such as Latina and Asian women (Noble, 2018: 82).

eating disorders got worse after using the platform (Romo, 2021). Haugen alleged that these outcomes were tied to Facebook's tendency to consistently prioritise greater profit, with little regard for how this could impact the well-being of its users.

It is not unusual for companies to be under pressure to maximise, even in ways that could be harmful to users. This is seen, for example, in debates over the promotion of tobacco and the fast-food industry. Still, these insights suggest that social media companies might be pursuing profit at all costs, including in ways that reinforce harmful norms for children and youth. Furthermore, when companies hide these algorithmic processes, users are not fully aware of how their information environment is potentially reinforcing particular norms.

Some studies suggest that profit incentives drive algorithmic outcomes in terms of what content is amplified. Roberts (2018) argues that concerns over profit underlie gender stereotypes, with algorithms amplifying content that promotes 'traditional' or patriarchal views of the female body. On social media platforms, sexualised images of the female body are seen to be more marketable (see Federici, 2004). Similarly, Bishop (2018: 81) finds that YouTube's algorithm rewards content from female users that promotes hegemonic (i.e. dominant) types of feminised cultural output, defined as consumption, fashion, baking and beauty. YouTube recommends content that aligns with social beauty norms because it aligns with brands, and advertisers' views of what is marketable.

Similar processes are visible on other social media platforms, such as dating apps (Gieseking, 2017), and for beauty ideals among different groups. Tyler Quick's (2021) ethnographic study of homoerotic content on Instagram reveals how social and algorithmic biases compound one another, and lead to the overrepresentation of images of white, predominantly American accounts as desirable homoerotic content. This informs algorithmic content moderation and means that gay Instagram users are increasingly shown white homoerotic content. In privileging white homoeroticism, users provide information to algorithms on their sexual preferences. This can cultivate online beauty standards that can leave intact and reproduce a hierarchical sexual culture (Green, 2013). According to Bishop (2018), the net result is a gendered splitting of content on social media, in terms of what is most popular and most visible to men and to women, in ways that reflect hegemonic views of beauty, femininity and masculinity.

Algorithmic outputs on social media platforms have also shown discrepancies in which content is targeted to whom. In 2016, it was revealed that Facebook enabled advertisers to target job and housing advertisements towards people with specific demographic characteristics (i.e. race and gender). This type of discrimination is illegal under US law. The practice was also criticised for putting many users in potential danger, particularly gay people living in countries where homosexuality is criminalised (Stokel-Walker, 2019).

Since then, demographic targeting has been restricted. In March 2019, Facebook disabled demographic targeting for housing, credit and job advertisements. However, ongoing research indicates that bias still exists. For instance, targeted content can differ for men and women. An audit by researchers at the University of Southern California revealed that Facebook's advertising system showed different jobs – but with the same qualification requirements – to women and men. Advertisements were targeted to reflect existing demographic distributions in jobs between men and women, thereby reinforcing existing disparities (Hao, 2021).

Finally, algorithmic outputs can negatively impact users by making different content visible depending on their sex and gender identification. Algorithmic content moderation means that algorithms play a role in deciding what user information is made visible to other users. Therefore, visibility is not entirely the choice of the user. To illustrate, while Facebook offers different privacy settings to its users, its default settings and underlying values promote greater sharing between users.[11] Facebook encourages users to connect to people they already know in other social spheres, including work, family and school, and default settings allow for someone's online activity, e.g. events they are attending, to be automatically publicised to the newsfeeds of users in their network (Duguay, 2016; Cho, 2018). Automated sharing of someone's activity can be potentially detrimental to those who might experience harm as a result of their identities being publicised. Cho (2018) shows how this 'default publicity' has not been a neutral terrain for queer youth of colour in the US. Cho found cases where information on users' sexual identities was broadcast to relations who were not aware.[12]

## Limits of automated outcomes

Algorithmic outputs are not the end of the story when it comes to the construction of gender norms on social media platforms. What users say and do online also plays a role, as this is the reference point from which data is created and labelled, which informs algorithmic decision-making. The algorithmic outcomes discussed in the previous section do not reflect user experiences globally. The ways that users choose to engage on social media platforms can lead to experiences that differ from some of the more dominant patterns in outputs that have been identified.

Social media users can, and do, circumvent the expectations and intentions of social media companies for platform operations. Looking at Facebook use in the town of Mardin in Turkey, Costa (2018) argues that users can choose how to respond to platform design and subvert Facebook's 'default publicness'. Costa shows how users created multiple profiles to communicate with different groups of people, also using fake names and pseudonyms. They were able to keep their social spheres separate in ways that would not have been possible if they had abided by Facebook's rule of one person, one user account.

In other instances, users who identify in non-heteronormative ways utilise social media platforms to connect with like-minded users in semi-public forums. This stands in contrast with the hegemonic gender associations identified in the previous section. In some cases, this has provided individuals with access to safe spaces and supportive communities online, whereas their identities might be criminalised or stigmatised offline. Bryan (2019) shows how sex and dating apps for men interested in men have provided access to communities and mechanisms for survival (e.g. facilitating sex work) in Uganda, where homosexuality is criminalised. LGBTQI+ Ugandans have used social media sites to navigate dating and work, and even to arrange 'lavender marriages' between gay men and lesbian women as a tactical survival mechanism (Bryan, 2019; see also Tamale, 2003). In South Africa, as Andrews (2021) explains, LGBTQI+

11 The introduction to Facebook's community standards states: 'Every day, people use Facebook to share their experiences, connect with friends and family, and build communities. It's a service for more than two billion people to freely express themselves across countries and cultures and in dozens of languages' (Meta, n.d.a).

12 Cho (2018) contrasts Facebook's design of 'default publicness' with Tumblr. Tumblr, at the time of Cho's research, did not require each individual to have one user identity, and the platform evaded easy indexing. Cho found that queer youth of colour in the US preferred to use Tumblr, rather than Facebook, to express intimate feelings.

vloggers and viewers have used YouTube as an arena to express themselves in more authentic and safer ways than would be possible offline. Lovelock (2017) shows similarly how YouTube influencers have used the platform to publicly express pride in diverse sexual identities. These online communities, and relations between influencers and followers, have become spaces of emotional labour, care and advocacy (Abidin, 2019). These forms of agency and community-building show that it is possible for users to share diverse content in specific spaces on social media platforms, presenting different perspectives and experiences of gender and sexuality.

In summary, both algorithmic outputs and user activity play a role in what becomes visible on social media platforms in ways that simultaneously constrain and support different gender identities. On one side, algorithmic outputs have tended to disproportionately amplify images of white people with perceived normative genders. Automated algorithmic processes also take some control from users in deciding who sees what information about them. On the other side, users make choices about how to navigate social media platforms, circumventing their intended designs and contributing to algorithmic processes. As a result, platforms are also spaces where some users can safely curate public appearances and connect on their own terms with particular communities or like-minded groups.

## 3.2 Data and algorithms as a form of disciplining power

### Discipline through digital data

So far, this chapter has looked at how social media infrastructure shapes the content that appears to users on the front end of the platform. Company priorities and profit, the platforms' technical design and users' activities interact to give rise to different expressions of gender norms, some dominant and others hidden among specific groups online. The back ends of platforms – where data labelling and machine learning algorithms take place – have also been found to shape gender norms, in and of themselves. This section explains how, and with what effect. Gender norms materialise in two key ways at the back end of social media platforms: (1) in the reduction of gender to discrete data categories, removing/ignoring other ways of defining gender; and (2) in the ways this data is used to continually shape and reshape the infrastructure for communication between people.

Cheney-Lippold (2011) argues that back-end data-based processes act as a form of 'disciplining power' because they condition how people identify and behave. This manifests through algorithms making inferences about user preferences based on data processed from users' behaviours online. This process of conditioning behaviour has two parts: (1) the construction of data, and how it relates to people; and (2) algorithmic decision-making on the back of this data.

First, users' ongoing activity online provides a constant stream of data for algorithms to process and 'learn from', so that they can update how they target a particular user. The structure of data, discussed in Chapter 2, limits how gender can be conceptualised. Data presents information as a string of discrete values; this string can then be processed by algorithms to identify patterns and make decisions. Data must be labelled, in other words given an 'identity', in order for an algorithm to filter information from that data (Apprich et al., 2018). Labelling involves attempts to make order out of messy data, and involves

making judgements (West, 2020). No matter how complex the algorithms or how granular the data, digital data is always constrained by the need to represent information as a sequence of discrete values. This means that gender is represented and labelled as discrete categories – that is, categories detached from, for example, societal contexts.

Second, algorithms make 'predictions' about people's immediate and future preferences based on patterns identified from accumulated data trails. Arvidsson (2016) argues that processing data created alongside people's past behaviours online, in an attempt to influence people's future choices, is inherently oriented towards reproducing what is already familiar, rather than towards imagining new possibilities. O'Neil (2016) argues that machine learning systems amplify past forms of prejudice, in an ongoing feedback loop that starts from historical data and existing practices.

Because of existing power relations, along racial, gendered and other historical lines of oppression, machine learning algorithms reflect and amplify existing inequalities (see e.g. West, 2020). Algorithms rely on data from human behaviour, which is produced within, and reflects, racialised systems of oppression. Biases emerge through different aspects of technological design and operation, for example, through humans labelling data or designing the original sets of rules (Benjamin, 2019). Benjamin (2019) argues that while claims are often made that digital technologies are relatively less biased than previous systems, this ignores bias in the process of designing technologies, starting with the designers and the context in which they exist.

To avoid reproducing these biases, data scientists must be explicitly cognisant of these dimensions of power, which are entangled with the design and running of algorithmic systems and how we experience them (West, 2020). This involves recognising and challenging power dynamics that frame how we work with data and algorithms (D'Ignazio and Klein, 2020). D'Ignazio and Klein (2020) explain and explore how data feminism is critical for actively countering power imbalances in the world and the complex ways they are reflected and amplified in data-driven technologies.

Chun (2011) argues that social media users, because they lack knowledge of computer programming activity, often approach these platforms with a degree of ignorance and ambiguity. Machine learning processes make predictions about social media users based on their past data, but these processes are not fully known to the individuals concerned. The more complex that algorithmic systems get, the less transparent they can become. Therefore, algorithmic systems are both predictable in one sense, in that they identify patterns from past behaviours based on precise instructions, and unpredictable, in that their operation is often opaque by being hidden from, and ambiguous to, human interpretation.

These complex algorithmic processes have resulted in the misgendering of users, meaning that algorithms can incorrectly assess a person's gender and then structure their online environment on this basis. Fosch-Villaronga et al. (2021), using a non-representative sample of Twitter users, found that misgendering is more common for gay men and straight women relative to straight men. Misgendering raises questions not only about algorithmic accuracy, but more fundamentally about whose identity-based characteristics the system can and should attempt to infer beyond users' own self-presentation.[13] When users are misgendered by

13 While this report does not include similar studies of misgendering on Facebook, where gender is a mandatory field for users, this remains a possibility, especially given that Facebook's advertiser platform consolidates gender into three categories and therefore misaligns with the 58 options available to users.

algorithms, this implies that something is not 'correct' about their appearance and engagement online, and, indeed, that they are misgendering themselves (Fosch-Villaronga et al., 2021).[14]

The technical response to misgendering has been to try to retrain or improve the algorithmic rules to better align with how people might identify. Software engineers have attempted to address inaccuracies by making data more precise and granular, adjusting algorithmic rules and/or training algorithms on expanded datasets.

However, altering the characteristics of data or algorithms does not remove the potential disciplining power of algorithmic systems. The technical response operates on an assumption that gender can be organised into discrete categories, contrasting with definitions of gender as more fluid and dynamic (e.g. Burgess et al., 2016). It also can raise new concerns about data harvesting, and about who is given access to increasingly precise data on individuals (D'Ignazio and Klein, 2020: 32). D'Ignazio and Klein (2020) argue that in most cases better detection, in this case of people of colour, ends up being used for targeted surveillance and harm. When considering unequal power relations – whether around race, gender or other characteristics – technical solutions are insufficient as they do not address the wider inequalities and biases that underpinned technical inaccuracies in the first place. Nakamura (2002) argues that identity presented as a menu of options is a form of stereotyping. From this perspective, misgendering is potentially an inevitable feature of a system that attempts to represent gender in discrete categories; there will always be opportunities for misalignment when translating a dynamic and fluid identity into digital data.

This possibility raises bigger questions about the accuracy of targeted advertising on social media platforms. There is little research on accuracy, likely linked to the opacity of platform algorithmic systems. Research that does exist suggests that targeting around different identity-based characteristics has had variable success, in line with evidence of instances of misgendering. Sances (2021) conducted a series of surveys between 2016 and 2018 to assess the accuracy of targeted ads on Facebook and found that accuracy varies.[15] For example, only 24.0% of ads were successfully targeted to African Americans, compared with a 99.8% success when targeting by age. Focused on the US context, and not looking at sex or gender, this study indicates the value of further inquiry into the extent to which targeting unfolds as intended, as well as the subsequent implications for gender.

The fact that targeting may not be accurate is not necessarily surprising. Gilroy (1993) argues that racial and gender categories are not fixed, meaning that any attempt to target someone by their perceived race or gender is likely to contrast with the potential dynamism in how they identify.

---

**14** Their study only looked at English-language users, leaving questions open about whether and how misgendering might occur in other languages, e.g. depending on how gender is embedded in the language itself.

**15** This took place prior to restrictions on targeting along age, ethnic and racial differences.

Facebook's community standards assert that Facebook is guided by authenticity, with each profile representing one person. This claim, combined with practices that simplify and categorise gender, projects a view of authenticity that assumes gender is sorted into specific categories.[16] Facebook's construction of authenticity proscribes terms that exclude many fluid, multifaceted identities, in particular by compelling people to use the name they use in 'real life' on the platform. This paradoxically prevents authentic self-representation for those who might not identify with their given name (Haimson and Hoffmann, 2016). When gender is taken to be fluid and dynamic, the problem of misgendering goes back to the very practice of representing gender as data points.

## Discipline through shadowbanning

Another way that algorithmic processes might intentionally discipline people's behaviour is through the practice of **shadowbanning**, whereby 'content moderators block or partially block content in a way that is not apparent to the user or their followers' (Bridges, 2021). Shadowbanning means that a platform, in practice, 'chooses only to connect some bodies' (Bridges, 2021). While the use of shadowbanning indicates that specific targeting takes place in platform content moderation, platforms do not disclose the relative roles played by human decision-making and human-programmed algorithmic decision-making in the practice.

Organised opposition to specific gender equality-focused groups and feminist activists can lead to gender equality content being shut down, as Hanna Paranta, whose Facebook account was disabled multiple times, experienced. Although her account was only shut down temporarily, Paranta's content has not returned to the status it had prior to being shut down. In particular, Facebook has not verified her page as the authentic account of a public person, which is done with a blue verification badge (Mahmood, 2021).

Anecdotal experiences reveal that shadowbanning takes place on different platforms. Are (2020; 2021) has investigated instances of shadowbanning aimed at adult performers, activists, athletes and people of colour. On Twitter, this means hiding user accounts from search results (Are, 2021). Instagram users have found themselves excluded from Instagram's 'Explore' page, where users are recommended new content. TikTok has been accused of hiding or deprioritising content by Black creators on news feeds. Claims about shadowbanning come from different groups, including political conservatives.

Shadowbanning provokes tensions between users and social media companies. Some users use humour and lobbying efforts to draw attention to shadowbanning. For example, sex worker and social media influencer @Charlieshe uses playful posts to speak to content moderators/platforms about shadowbanning practices (Bridges, 2021). Instagram in particular has been ambivalent in responding to user concerns. Instagram denied shadowbanning took place until mid-2019, when a petition by almost 20,000 pole dancers led to an official apology from Instagram to the account bloggeronpole.com. However, in February 2020, Instagram's CEO Adam Mosseri again denied that shadowbanning took place. This was contradicted in June, when he published a blog post that acknowledged that it does (Are, 2021).

16 Lingel and Golub's (2015) study of Brooklyn's drag community and its battle with Facebook's 'real name policy' highlights the platform's design incompatibility with diverse approaches to gender performances and naming practices.

Shadowbanning provides a reminder that people, and not just algorithms, play a role in algorithmic content moderation, whether in designing or adapting algorithmic processes, or in overriding their results. Here, the people within a social media company make judgements that shape what content appears or is demoted. These decisions are often hidden from users, which makes it difficult to identify whether this is taking place through algorithmic programming or human intervention. This indicates that social media companies use a degree of obscured censorship to decide how people should appear in public. These practices have in some cases targeted women and feminist activists.

## 3.3 How users reify gender norms in an attention-based industry

### Influencers and the gendered nature of work

This section looks at how users have become involved in profit-making on social media platforms, and how this too contributes to reifying certain gender norms. Some social media users take part in the targeted advertising business as **influencers**. By 2019, influencer marketing was estimated to be a $9 billion industry (Bertaglia et al., 2020). The influencer industry partially emerged as a response to concerns that third-party targeted advertising on social media had limited effectiveness. Users could see it as intrusive or disruptive, and could block advertisements through ad-blocking software (De Veirman et al., 2017). Influencer-based advertising responds to these limitations, by having popular users advertise products to their networks. Brand or marketing firms pay these influencers to promote commodities to their audiences. An influencer's commercial value is tied to their visibility, and their perceived authenticity and credibility (Cotter, 2019). Influencers weave marketing into their personal content, treading a fine line between building up and maintaining the loyalty of followers and selling products on behalf of advertisers (Van Driel and Dumitrica, 2021).

Influencers contribute to gender norms in multiple, sometimes conflicting, ways. First, influencers, especially in Anglo-American contexts, have tended to reinforce stereotypical images of women's work and femininity. Influencer work is most often associated with women and industries traditionally seen as female (e.g. beauty, fashion, crafts, parenting and homemaking) (Van Driel and Dumitrica, 2021). Their content also has been found to promote certain representations of women. Duffy and Hund (2015) identify three interrelated tropes that emerge among top-ranked fashion bloggers: the idea that passionate work is destined to be successful, the staging of a glamorous life and carefully curated social sharing. There also are some indications of the bifurcation of male and female influencers on different platforms. There are indications that most influencers with sponsored posts on Instagram are women (Statista, 2021b). Bishop (2018), in a sample of popular vlogger accounts, found that the most subscribed-to independent vlogger accounts in the UK were mainly owned by men (43 out of 50), with popular men's accounts covering topics such as gaming, sports, technology, comedy and news. Popular women's accounts covered topics such as beauty and children, although one focused on gaming.

Second, influencers' entrepreneurial activity is linked to individuality and consumption, which contribute to particular images of feminism and women. Women lifestyle influencers use physical capital (their appearance) and connections as resources for financial independence. Feminism becomes an individual choice rather than a collective struggle. In marketing products, women's entrepreneurship is linked to consumption (Petersson

McIntyre, 2021). In some cases, feminism itself is used to attract followers, contributing to a commercialised view of feminism (Mahoney, 2020). These patterns suggest an influencer landscape that is dominated by women, and which reflects specific material and stereotypical images of women.

While the overarching picture of the influencer suggests a tendency towards stereotypical gender displays, influencer content varies. Chen and Kanai (2021) show how four gay male influencers gained popularity as beauty influencers, an area of marketing predominantly occupied by women. Viral content can challenge gender binaries. In '#TheBoyChallenge' on TikTok, users created and posted short videos in which they presented a simple gender transformation. Their videos showed that gender is a performance that can be transformed (Khattab, 2019). Other influencers, such as some LGBTQI+ influencers on YouTube, demonstrate their authenticity through showing pride in their sexuality (Andrews, 2021). Lovelock (2017) studies two celebrity YouTube influencers who identify as lesbian and gay, and shows how their posts, in which they explore their journeys from shame to self-acceptance and pride, help to strengthen public acceptance of their sexual identities.

Evidence about the influencer industry is limited. It is mainly informed by research in Anglo-American contexts and focuses on women influencers. Influencers exist outside Anglo-American contexts and have varied relationships to gender norms. Wilson et al. (2018) examine the account of a Malaysian female influencer, under the username @sareesandstories, who provides narrative storytelling around posts, usually of saris, about the identity construction of Indian women. Ligaga (2016) draws attention to the presence of women 'socialites' in Kenya, who self-represent and maintain celebrity status through Instagram/social media accounts. Their celebrity status is tied to different forms of income generation; for example, rather than posting products they charge for their participation at events (e.g. Kenyan influencer Vera Sidika has claimed to charge more than 280,000 Kenyan shillings, ~US$2,500, an hour). Ligaga (2016) argues that Kenyan socialites on Instagram challenge patriarchal social norms, even while potentially reflecting gender stereotypes and social inequalities.[17]

## The role of infrastructure: how influencers navigate marketability and algorithms

Influencers play an active role in shaping what content captures attention. However, their influence is precarious. They need to remain profitable to third parties, popular with their audiences and visible within platforms' technical systems (Bishop, 2021). Profitable influencers must navigate between authenticity and monetisation, self-commodification and objectification. There are high stakes for failing to tread this line. Banet-Weiser (2021a: 143) suggests:

> Authenticity on social media, then, is framed by a profound tension: for female influencers on Instagram, being authentic is often about constantly adjusting yourself to correspond with dominant white ideals of femininity. Yet authenticity is also about failure, pressure, depression, tears, vulnerability. This is the labour of authenticity.

17 In another example, focusing on sexualised content of 172 female influencers on Instagram, Drenten et al. (2020) found that heteronormative ideas of attractiveness and femininity shape how influencers garner attention.

Looking at a sample of female online entrepreneurs, the majority of whom were white and college-educated, Duffy and Pruchniewska (2017) conclude that these influencers are caught in a 'digital double bind' where they must operate within three imperatives: a subtle, soft approach to self-promotion, an interactive intimacy with audiences and a compulsory visibility requiring them to make their private lives public.

The extent to which these expectations frame other groups of influencers is a question for further study. Research seems to be disproportionately oriented towards Anglo-American contexts, and focuses on either female or male influencers, with less attention to men and to the potential differences between and within gender groups. Vaiciukynaite (2019) indicates this as an important area for further research, noting some indication of differences in the experiences of male and female influencers on Instagram. While finding little difference in how men and women respond to female influencers, Vaiciukynaite found differences in how they respond to male influencers: female users spent more time viewing their content, while male users were more likely to click 'like' in response.

A 2019 marketing study has also identified discrepancies in pay. Surveying over 2,500 influencers on Instagram, YouTube and Facebook, the study found more female influencers than male (women made up 77% of the sample). However, on average, women charged less than men (US$351 and US$459 per post, respectively) (Young, 2019). Influencers' profitability depends on engaging in ways that are appealing both to other users and to marketing firms. What this entails can differ for men and women, although exactly how is a question for further research.

Alongside users and marketing firms, platforms also constrain how influencers can engage to retain their visibility and, therefore, profitability. Influencers need to be visible on these platforms, which utilise algorithms to amplify and demote content. Often, influencers must infer how platform algorithms amplify content; given that they are usually proprietary, influencers have limited information about the algorithmic rules that shape their visibility (O'Meara, 2019). Cotter (2019) suggests that algorithms encourage influencers to behave in certain ways to remain visible. However, influencers have also learnt how to engage more strategically with algorithmic rules. For instance, in 2016, Instagram influencers adopted several strategies to adjust to the platform's switch from presenting content chronologically to a preference-driven feed linked to users' usage history and the popularity of their posts. Some influencers formed a collective organisation and worked together to improve their engagement rates (O'Meara, 2019). Some started a change.org petition to ask Instagram to reconsider its strategy, which gained more than 340,000 signatures. Another response was to form 'engagement pods' – groups of users who agree to mutually like, comment and share each other's posts, as well as share insights on how to effectively work with the platform's algorithms.

Influencers enable users to actively shape the content that others see on social media platforms. However, their influence is bound up with a complex set of relationships and dependencies, in which they must respond to profit incentives and platform design.

## Tumblr as an alternative technical (and unprofitable?) model for user activity

Tumblr has some distinct technological design and business features that differ from the more dominant and profitable social media platforms, such as Instagram, Twitter and Facebook. It offers an important study of gender norms online, as it has been said to allow more flexible and open use by LGBTQI+ communities. Some scholars describe Tumblr as a queer space as it has enabled queer communities to connect and embrace fluidity in how people present their identity (Fink and Miller, 2014; Cavalcante, 2019). Haimson et al. (2021: 357) suggest:

> Tumblr is a fascinating case study: a site somewhat accidentally designed with queer and trans features, legible to and used by queer and trans people, that temporarily existed within the capitalist framework of Silicon Valley.

Specific features contribute to this image of Tumbler. Tumblr's origin narratives suggest its founder, David Karp, wanted to create a space for short multimedia expression. From here, it then quickly became seen by users as a site that was devoted to self-expression and permissiveness, and it attracted a mix of users (Cavalcante, 2019). Technically, Tumblr allows users to remain anonymous and displays content chronologically. Contrasting with Facebook, where an embodied individual signifies authenticity, Tumblr does not emphasise connections between online identities and wider offline connections (Cho, 2018). Users can choose to integrate different media into their pages to express themselves (Cavalcante, 2019).

For trans* individuals, Haimson et al. (2021) find that anonymity and separation from offline networks make Tumblr a safer space to express themselves than offline spaces and other social media platforms. This was important for some as they were changing and experimenting with appearance. The trans* users interviewed explained that Tumblr allowed them to self-present as the person that they were 'becoming', whereas Facebook was designed to present a user as a fixed entity that aligns with one gender assignment. Tumblr was described by some as a trans* technology because it provided for the multiplicity and fluidity of identity, and offered separation from offline networks.

Still, Tumblr's features have had ambivalent effects. Haimson et al. (2021) suggest that this is partially because Tumblr was not designed with specific (e.g. trans* or queer) users in mind. Focusing on trans* users, they show how these users have had to work to protect their engagement online, for example in response to Tumblr tagging their content as 'adult' or reacting to porn blogs following their accounts. Tumblr's permissiveness also allows for racist and homophobic content, and homophobic Tumblr communities exist alongside LGBTQI+ ones (Cavalcante, 2019; Haimson et al., 2021). Cavalcante (2019) also argues that queer communities on Tumblr have tended towards an echo chamber effect, as users increasingly engage with like-minded people and limited content, reinforced by platform recommendations that encourage users to follow other bloggers based on their existing network.

The value of a safe space for expression online and the potential for harm increases for people with intersecting marginalised identities. Trans* people of colour have experienced intersecting patterns of racism and transphobia, which has amplified their marginalisation. At the same time, bloggers have also identified opportunities for giving and receiving support from others on the platform, working to create sub-groups within trans* communities. However, they noted that this requires additional work and resources (Haimson et al., 2021).

Finally, Tumblr has struggled to sustain the same profitability and growth as other social media platforms, with more technically complex processes for targeting content. As mentioned in Section 2.1, Tumblr fell in value from $1.1 billion to $3 million from 2013 to 2019 (Siegel, 2019). Still, this case is important in that it shows the possibility of alternative ways in which business, technical design and user activity can come together to provide for different – albeit still ambivalent – experiences and expressions of gender norms.

## 3.4 Community guidelines: company policies and their consequences for gender norms

Beyond what algorithms or users do, most social media companies set out community standards or guidelines that lay out a normative framework about what they consider to be problematic content and how it is dealt with. Community standards or guidelines are written policies from the company behind the platform about how they identify and address problematic activity and content. Policies cover content and user behaviour (e.g. harassment), and set the terms from which platforms moderate and censor content. Community standards vary in details and length (Bateman et al., 2021).

Standards are also dynamic. Facebook updated 21 out of its 27 policies at least once between September 2020 and February 2021. Moderation of content has evolved as user bases have grown, changing the scale and scope of community values. Companies often change their community guidelines in response to community requests, sometimes on the back of controversies. For instance, following public outrage, Facebook invested in improving its Myanmar language hate speech classifiers, which resulted in a 39% increase in posts being taken down in the following six-month period (Gorwa, et al., 2020). In October 2019, Facebook also created an independent Oversight Board to help set precedents for, and make decisions about, content in line with its community standards (see Box 3).

Gillespie (2017) suggests that, over time, a degree of homogeneity has emerged among platforms' standards, with most now prohibiting sexual and obscene content, violent content, user harassment, hate speech, promotion or representation of self-harm, and promotion or representation of illegal activity. Gerrard and Thornham (2020) argue that this stability in community guidelines mainly pertains to areas where content would be deemed to be illegal, including support for terrorism, crime and hate groups, and sexual content involving minors. Greater dynamism remains around other issues, where existing laws do not necessarily clearly apply. Platforms' user codes of conduct also vary. For example, Twitter prohibits coordinated abuse through technical means, whereas YouTube employs a 'three strikes' system, from warnings to eventual channel termination (Jankowicz, et al., 2021: 10).

### Box 3: Case study – Facebook's Oversight Board

Facebook has introduced a distinct model of self-regulation in the form of an independent Oversight Board launched in October 2019. The formation of the board was informed by a series of workshops held globally. A public portal was created for global recruitment to the board. The board is 50% female, with representation from different areas of expertise, global locations and ethnicities. In reporting on the process, the *New Yorker* revealed that Facebook's CEO supported the uptake of the idea, enabling it to be taken forward irrespective of internal dissent.

The board uses Facebook's community standards as the reference point for making decisions. Facebook has one set of community standards, intended to reflect the platform's global user base and the diversity of contexts in which it operates. The board can make decisions on cases of content moderation and make recommendations to Facebook about its content policies. Cases can be referred to the board by Facebook or by user appeals, following a decision by Facebook to take down or leave up content. While the board does not have legal or enforcement authority, Facebook states that its decisions about content removal cases are binding.

So far, little external research has been done on the Oversight Board's approach to regulating content. Some scholars have questioned how Facebook's singular set of community standards impacts a diverse user base; for example, Spišák et al. (2021) highlight how users engage with a range of identities and sexualities, and with interests that do not necessarily align with Facebook's community standards (see also Tiidenberg and van der Nagel, 2020). The standards can therefore contribute to stigmatising practices and cultures that exist outside one set of ideals. The Oversight Board might help to mediate between different interests, through a globally diverse membership of external experts and civic leaders, but still aims to come up with one agreed outcome.

*Sources: Meta (n.d.a); Oversight Board (n.d.); Oversight Board (2020); Klonick (2021).*

## Content removal and the enforcement of community guidelines

Community guidelines set the parameters for content removal. Content that is determined to be problematic can be subject to moderation in specific ways; for example, content can be restricted in its accessibility or searchability, or it can be removed. Also, in extreme cases, a user can be temporarily or permanently removed. Removal prevents further or wider offence but can conflict with principles of open participation and protected speech. Also, it is not necessarily fully effective. Content removal can result in a game of 'whack-a-mole' where users create new accounts and continue to post. This game indicates the challenges of labelling any content as problematic, from feminist campaigns that might be labelled as misinformation by anti-gender groups to hate speech towards women (see e.g. Banet-Weiser, 2021b). Filtering can appear less invasive but can be relatively invisible to the user, hiding the scope and nature of moderation (Gillespie, 2017).

Computational tools and human monitoring are both involved in upholding community guidelines (Gillespie, 2018). Most platforms use a tiered system of content moderation to deal with problematic content. This enables them to cope with the scale and speed of content creation and sharing. Internally, companies often have moderation teams that work with frontline reviewers and independent contractors.[18] External to the company, content removal can involve crowdsourcing platforms, with workers hired to review content

18 For example, Amazon Mechanical Turk, a crowdsourcing site that allows businesses to hire remote workers for discrete tasks.

on a per-task basis, volunteer moderators, as well as users who voluntarily flag content as inappropriate. Most platforms allow users to report or 'flag' offensive content to the platform and trigger a review process (Crawford and Gillespie, 2016; Guo and Johnson, 2020). Flagging is optional (Gillespie, 2017).[19] Crawford and Gillespie (2016) suggest that flagging both enables and obscures collective governance. Users can participate in governance by labelling content as problematic – in contrast to the wider activity of data labelling in which users are labelled by others. However, the decisions made on the back of users' 'flags' are not visible to the individual or to other users. Furthermore, while flagging enables users to participate in content moderation, it does so in a way that often obscures their labour, which arguably is a form of care labour that is insufficiently remunerated (The Good Robot, 2021).

Social media companies cannot avoid using automated content moderation, given the scale of social media activity globally. External regulations – for example, Germany's Network Enforcement Act or the European Union's (EU) Code of Conduct on Hate Speech – can also impose restricted timeframes for identifying and removing content that necessitate automated removal (Gorwa et al., 2020).

Large tech companies, including some social media companies, have cooperated in algorithmic content moderation around key issues, such as counterterrorism. Facebook, Google, Twitter and Microsoft formed the Global Internet Forum to Counter Terrorism in 2017, and worked together on a common hash database of terrorist content.[20] Alongside this, platforms retain their own algorithmic architecture for content identification and removal. For example, Facebook has begun to use its own machine learning algorithms to create a predictive score for how likely a post is to violate Facebook's terrorism policies (Gorwa et al., 2020).

## Gender bias through the enforcement of community standards

The creation and enforcement of standards for content removal expose a normative assessment of how people should present themselves and interact. Southerton et al. (2021) suggest that content classification systems turn platforms into 'norm-producing technologies', as they decide and enforce views of problematic content. Norms that materialise through the removal of content can reflect either explicit or hidden biases. They can even contradict platforms' written standards that value dignity and rights. For example, so-called 'unintended bias' along racist or sexist lines has been found in training datasets that are used to train machine learning algorithms to match or classify content. This was visible in the case of Perspective, a project by Jigsaw, a Google/Alphabet subsidiary, which predicts the impact of a comment on a conversation. Data for Perspective was trained by workers from a crowdsourcing site, and revealed bias along racial and political lines. For example, content including the word 'Arabs' was labelled as more toxic than content indicating Nazi sympathies (Gorwa et al., 2020).

19 Another form of user moderation is volunteer community management. This is visible at a site level (on Wikipedia, for instance) or in social media platform features such as Facebook groups or 'subreddits' (on Reddit), where individuals can manage content within groups that they administer.

20 This forum was part of their commitment to the European Commission's code of conduct to combat illegal online hate speech.

Past examples have revealed the uneven effects of content removal for men and women, and for LGBTQI+ users – both in what is removed and in what is not. First, content that expresses hate and/or harm towards women continues to be visible on platforms. This indicates that community standards are not enforced in ways that fully protect women from harm. A review of Facebook's internal documents in *The Guardian* found that comments that threatened women with violence were not removed as threats. This included comments such as 'To snap a bitch's neck, make sure to apply all your pressure to the middle of her throat', and 'Little girl needs to keep to herself before daddy breaks her face' (Nurik, 2019). On a bigger scale, examples such as the Rohingya genocide in Myanmar in 2018, where Facebook failed to monitor and remove content fuelling ethnic cleansing, indicate that the limitations of self-moderation might be even greater in languages other than English and in non-Western contexts (BBC News, 2018).

Different user groups have been disproportionately affected by content removal. Studies have focused in particular on the experiences of LGBTQI+ users. YouTube has been embroiled in a number of controversies in this regard. In February 2017, YouTube stated that it would strengthen its system for identifying offensive content, giving advertisers more control over what content is paired with their brands. This announcement took place after criticism that some of the videos it hosted included extremist and terrorist content. This decision had an unintended, negative impact on some LGBTQI+ creators on YouTube, who reported that their videos were hidden within a restricted mode (an opt-in setting from 2010). YouTube responded by allowing unrestricted videos featuring LGBTQI+ content (Southerton et al., 2021; see also Abidin, 2019). However, in August 2019, LGBTQI+ YouTube users filed a class action lawsuit against YouTube for discrimination, alleging that the platform's moderation of content, by both algorithms and people, continued to discriminate against LGBTQI+ content. Plaintiffs' cases included some users being told they could not have content promoted 'because of the gay thing' and blanket age restrictions on users who posted, for example, 'kink-friendly' sex education. As part of the case, some users explained that they had begun to self-censor content that could be labelled as queer by the platform's algorithm to avoid its removal, attempting to predict how its opaque interpretation of standards would be applied (Kleeman, 2019). These experiences of being censored contrast with instances where platforms have allowed content, including racist and homophobic content, to stay (Rosenberg, 2019).

Concerns over how community standards are enforced have emerged around other platforms too. Accounts can be removed in error; for example, a survey by Salty (2019), a member- and volunteer-driven digital newsletter and platform for women, trans* and non-binary people, found that half of its respondents' Instagram accounts had been reinstated after deletion. Tumblr has faced allegations that its content moderation and removal procedures unfairly discriminate against queer users. In 2018, Tumblr's suspension from the Apple App store, for hosting child pornography, triggered an internal decision to ban all **NSFW (Not Safe For Work)** content (Pilipets and Paasonen, 2020; see also Southerton et al., 2021). The ban affected NSFW blogs and fandom art, but many users criticised it as ineffective, as porn bots (fake followers and spam content) could evade the ban. Users contested the ban using the platform's technological features; for example, they tagged content to express criticism of Tumblr's policies, repackaged and reshared blocked content, and created clusters of content criticising the ban.

## Organisational and technical features that contribute to biased content removal

Sometimes content is unfairly removed due to the specific constraints of algorithms. In particular, algorithms struggle to take *context* into account when making decisions (see Khan, 2021: 81), operating instead from precise definitions. This can have significant effects, as disputes over content moderation decisions have shown. For instance, Facebook's algorithm for removing content based on female nudity did not distinguish photos of breastfeeding and female indigenous elders with uncovered breasts from other types of nudity (Dragiewicz et al., 2018). Spišák et al. (2021) suggest that it is relatively easy for an algorithm to identify nudity or sexual activity, but difficult to take into account its context. Unpacking patterns in Facebook's regulation of sexual content, Spišák et al. (2021) suggest that this occurs because Facebook has tended to focus on negative freedoms (e.g. protection from harm) rather than positive freedoms (e.g. the freedom of sexual expression). Content moderation policies focused on protecting people from harm contribute to sexual stigmatisation, particularly for those who are outside heteronormative 'ideals'.

Nurik (2019) argues that the strict application of community standards has unequal effects. Facebook's censorship rules about content that threatens protected categories (e.g. race, gender and religion) usually do not take into account humour or irony. They also do not protect *subgroups* of protected categories. For instance, a female driver is categorised as a driver rather than as a woman (Nurik, 2019; see also Angwin and Grassegger, 2017). Therefore, if a post contains a sexist threat to a woman who has been categorised as a driver, this will not necessarily be interpreted as a threat under the platform's community standards, because 'driver' is not a protected category.

One contributing factor is the relative scale of content compared to the resources required for content moderation. Facebook has been increasing its number of content moderators, for example, from 4,500 in May 2017 to 15,000 as of 2021 (Madrigal, 2018; Simonite, 2021). Yet, while the majority of Facebook users operate in languages other than English, it has faced strong critiques for its capacity to accurately and comprehensively monitor content in different languages. Little information is publicly available about who is employed as censors for social media platforms; the unequal demographic make-up of organisations suggests that this is an area for further research. In particular, there is a need for research that explores how the conditions and intersections of human and algorithmic moderation might shape differential outcomes (Nurik, 2019).[21]

> Facebook's algorithm for removing content based on female nudity did not distinguish photos of breastfeeding and female indigenous elders with uncovered breasts from other types of nudity.

21 Nurik (2019) interviewed members of the 'Men are Scum' movement, a social media movement involving a group of female comedians posting comments using the phrase 'men are scum'. Many of the comments were banned, leading to debates over censorship (see Gibbs, 2017).

## Contesting platform labels

Although outcomes of content removal processes have been experienced as unequal and biased, the channels available to users to contest how they are being implemented are limited (Nurik, 2019). When Salty surveyed its members on their experiences of content moderation on Instagram and Facebook, it found that, in most cases, users were not given a specific reason why their advertisements had been removed from Instagram. Instead, they were simply told they were in violation of community guidelines. The survey also found that there was little recourse open to users to appeal restrictions placed on their accounts or content (Salty, 2019).

This indicates a power imbalance when it comes to who decides and interprets what is appropriate content. Social media platforms' terms of use contribute to this power imbalance. Article 14 in Facebook's terms of use allows the platform to ban users on the basis of anything that violates the letter or spirit of the terms, or that creates a risk or possible legal exposure for Facebook. Twitter's terms give the platform unconditional power to remove or refuse to distribute content, terminate user accounts or reclaim usernames.[22]

Furthermore, cooperation between governments and social media platforms tips the balance of power away from users. Governments can pressure social media companies to remove content, for example, as evident with anti-terrorist policies (Leerssen, 2015). Social media platforms receive frequent government requests for user data; in the latter half of 2020, governments made over 191,000 such requests, with most coming from the US, followed by India, Germany, France, Brazil and the UK (Meta 2021b). Facebook states that it complies with requests that are legally valid and consistent with international human rights standards (Meta, 2021b). In the same period, 72.3% of requests were granted (Meta, n.d.b). Platforms can exercise some discretion in how they respond to government requests. For example, Twitter has removed tweets in response to government requests but has made clear what was being removed, while Google pulled out of China after the Chinese government requested that it filter search results by government dictates (Gillespie, 2018).

Rather than work with limited platform channels for feedback or complaints, users' most visible efforts to contest content removal have involved collective action. West (2017) examines different strands of collective action against Facebook's nudity policy, which removed images of female, but not male, toplessness. These campaigns ranged from those rallying around images of women breastfeeding, to images of breast cancer survivors' mastectomy scars, to the hashtag campaign '#FreeTheNipple'. Humorous and original viral campaigns, some involving influencers, provided leverage for the cause, although did not result in a clear change in Facebook's policies.

22 According to Leerssen (2015), limited liability for platforms, as exists in US law, could indirectly protect users by protecting platforms from being liable for the content that is shared. However, there is no guarantee that platforms will preserve users' freedom of expression.

## 3.5 Reflections

The interactions between social media platforms' technical design, business aims and user activity may, at first glance, seem to amplify a relatively narrow and stereotypical set of gender norms. There are clearly some dominant tendencies in terms of which gender norms platforms promote, based on what has advertising appeal, what can be recognised as data and by algorithms, and what contexts or subtleties are lost in content moderation. In some ways, users have also reinforced these patterns through their behaviour. For example, the influencer industry often reinforces stereotypical ideas about women's interests and roles.

However, these interactions are dynamic, which also makes it possible for users to expand and diversify gender norms online. Some influencers take pride in non-heteronormative sexualities, while viral campaigns have revealed the unequal impacts of content moderation for women and queer communities. Users have also created spaces for self-expression with like-minded individuals on different platforms, including YouTube, Tumblr and dating apps. A focus on dominant patterns and platform design can overlook the diverse ways users that interact on and with social media platforms to explore and contest gender norms in ways that both promote marginalised identities and challenge gendered injustice (Washington and Marcus, 2022).

The next chapter shifts its gaze to external factors that influence the dynamics between technology companies and their users. It looks at how external actors – specifically governments – have attempted to control what happens on social media platforms, examining how such efforts have shaped, or been shaped by, gender norms.


© Markus Esselmark/UN Women

# 4 Existing legal frameworks and alternative models to regulation

According to The Economist Intelligence Unit, most women have experienced or witnessed abuse online, despite platforms' content moderation processes (The Economist Intelligence Unit, 2021; see also Di Meco and Wilfore, 2021: 2). Without external regulation, social media platforms appear to be easily taken over by content marked by prejudice, hate speech and abuse.

Given the prevalence of harmful content, external actors have become increasingly concerned with platform governance.[23] Debates about how to regulate content online emerged in the late 1990s, at first focusing on online defamation and porn, and later expanding to concerns about online piracy and copyright infringement (Gillespie, 2018). From the 2010s onwards, calls for regulation have often focused on the rise of misinformation and hate speech, antitrust issues, the monopolisation of the sector by a few firms and data privacy (see Box 4).

Regulations can alter or limit how platforms operate at different points. For instance, they may focus on technical infrastructure (such as algorithmic systems), limit partnerships between platforms and companies, or set conditions for what types of content must be removed.

This chapter unpacks different approaches to external regulation, mapping the different aspects of social media infrastructure and activity that are being targeted. It draws out existing evidence of how these regulations impact gender norms on platforms, and highlights some key gaps in existing research on this topic.

## 4.1 Current laws

While social media platforms have a global reach, national jurisdictions play an important role in regulating platform activity and content.

Most large social media companies are corporate and legal entities in the US (Gillespie, 2018). Many debates about who is responsible for content on these platforms therefore use US liability law as a reference point. This dates back to Section 230 of the 1996 Communications Decency Act (CDA), which offers companies safe harbour from liability for harmful user content, as they are not deemed the legal publishers of this content (Gillespie, 2017; Nurik, 2019).[24] It also provides legal freedom for companies to intervene on users' behalf (Gillespie, 2018). This means that platforms are alleviated of responsibility for content both when they do and do not intervene (Crawford and Gillespie, 2016).

---

23 Gillespie (2018) identifies two main approaches to moderating social media content: governance *by* platforms and that *of* platforms. The latter is external and specifies companies' liability for the content that they host. The former consists of companies' initiatives to curate content and monitor activity on their platform.

24 Initially, the law made it illegal to provide obscene or indecent material to minors but this was deemed unconstitutional a year later.

US liability laws protect companies from being responsible for content created by their users, with some exceptions related to child pornography and intellectual property. This is based on a particular understanding of social media platforms as *hosts* rather than as *publishers* or *creators* of content, which means they are not considered to be accountable for what users post. Large technology firms have lobbied the US government extensively, with the top seven firms spending $64.9 million in 2020. Almost $20 million of this came from Facebook, which increased its lobbying efforts during the run-up to the elections, when there was significant public attention on issues around social media content management. Other technology firms' lobbying efforts have coalesced on issues such as privacy, competition and antitrust legislation (Romm, 2021).

As social media platforms have grown, expanding their activity and the scope of their content, critics in both political and academic circles have questioned their legal designation as hosts rather than as publishers of content (Wakabayashi, 2020). Social media platforms have multiple interfaces: they work through

### Box 4: Regulatory debates over data privacy, antitrust and advertising

Debates over external regulation have come to bear on social media platforms around data privacy, advertising and antitrust issues. These alter the terms of platform business models, e.g. targeting, partnerships and data use, indicating their potential to disrupt the interactions between gender norms and platform infrastructure. However these issues have yet to be studied in depth.

Data privacy concerns came to the fore in March 2018, when a newspaper investigation by Cadwalladr revealed that Cambridge Analytica had used Facebook data, gathered without consent from approximately 87 million users, to conduct psychographic voter-targeting. Facebook responded to the Cambridge Analytica scandal by disabling data access for some third parties and tightening its terms of service. Facebook is not the only platform to alter its APIs over time. Such changes have been made for different reasons (e.g. Twitter's changes have focused on issues related to larger, commercial use of its data). Commercial social media analytics services also continue to exist. A mixture of concerns about privacy and profit incentives shape what data is accessible and to whom (Bruns, 2019).

The EU's General Data Protection Regulation (GDPR) is one of the earliest and most comprehensive data protection frameworks globally. GDPR applies to the processing of personal data relating to individuals in the EU. It regulates standards of consent for different types of data and specifies individuals' rights to access their personal data. GDPR has contributed to a language of ownership over data. By contrast, the US has lagged behind in legislation. In May 2021, Democrat Senators reintroduced the Social Media Privacy Protection and Consumer Rights Act. If passed into law, this would give consumers more control over their personal data. It would also provide the Federal Trade Commission and state attorney generals greater capacity to enforce data privacy.

Commercial advertising is another area of external regulation that could alter the way social media platforms operate. This also has implications for influencer activity. The US Federal Trade Commission requires social media influencers to disclose receipt of any compensation they receive for posts. In Europe, sponsored content that is produced by influencers is regulated under the Unfair Commercial Practices Directive and the Audio-Visual Media Services Directive.[i] Specific European countries vary when it comes to additional requirements, e.g. the Netherlands focuses on self-regulation while some German courts suggest that every influencer post requires advertising disclosures.

*Sources: Cadwalladr and Graham-Harrison (2018); Isaak and Hanna (2018); Privacy International (2018); Bertaglia et al. (2020); Burns (2020); Kang and Frenkel (2018); Klobuchar (2021).*

i The former sets out when it is unfair not to disclose that content has been sponsored. The latter sets out requirements for certain types of advertising, e.g. that aimed at vulnerable groups.

partnerships and programmable interfaces, and involve interdependent relationships between users and algorithms. Crawford and Gillespie (2016) argue that this means that platforms not only host but also shape the content of others. When someone posts on a platform, their content becomes input for algorithms, which feeds back into news feeds, trend lists and other platform features.

Other countries' liability laws vary. Russia and states in the EU and South America provide companies with conditional liability that is contingent on platforms' knowledge of content and requires them to respond to state requests to remove illicit third-party content. China and countries in the Middle East tend to have stricter liability laws that require companies to prevent the circulation of unlawful or illicit content. Some countries have no liability laws for social media platforms (Gillespie, 2017).

## The regulation of harmful content

In addition to laws that regulate social media platforms' liability for content generally, there are more specific areas of concern that have become subject to regulation. This includes activities that are specific to social media platforms as well as general issues around digital technology and data.

Again, US legislation remains a key reference point. Since Section 230 of the CDA was introduced, the US government has placed limits on safe harbour provisions in relation to illegal content. In 2018, two acts were passed – the Allow States and Victims to Fight Online Sex Trafficking Act (FOSTA) and the Stop Enabling Sex Traffickers Act (SESTA) – that introduce exceptions to Section 230 of the CDA. In response to these acts, Facebook tightened its community standards and terms of use (see Box 4 for different areas of regulatory debate).

Other countries have taken different approaches to moderating sexually explicit content. Some governments, including Germany, France and Austria, have conducted audits on specific areas of harmful content, such as hate speech. In 2018, Germany instituted the Network Enforcement Act, which requires social media companies to respond to illegal content flagged by users within a specific timeframe. The UK government's 'Online harms' white paper proposed a 'duty of care' approach, which demands greater transparency from platforms (Department for Digital, Culture, Media & Sport and Home Office, 2019). The EU Digital Services Act places due diligence obligations on large platforms and requires them to conduct regular risk assessments.

While much research on content removal has focused on what platforms do, platforms' community standards do not exist in isolation, and platforms must contend with legal requirements in the jurisdictions in which they operate. McLelland (2016) suggests that external regulation can constrain how different sexualities are viewed and experienced on platforms. Comparing the censorship of obscenity in China and Japan, McLelland (2016) argues that national regulations on obscene or pornographic content can restrict speech and criminalise some forms of sexual expression. Japan's approach to obscenity distinguishes between fictional characters and real people. This has given people relative freedom to express themselves in diverse ways through fictional characters. China has placed greater restrictions on obscene content, limiting what can legally appear on social media. For example, since April 2014, the Chinese government has imposed a 'Cleaning the Web' campaign to remove 'pornographic and vulgar information', which it argues harms the health of minors and 'seriously corrupts social ethos' (McLelland, 2016: 124).

## 4.2 Alternative models of external regulation

### Algorithmic impact assessments

There is increasing interest in platform algorithms and the roles they play in determining the visibility and removal of content. Yet, algorithmic activity remains a relatively new area for regulation and is a topic that requires further research. Research could help to identify what technical understanding and skills are needed, as well as to critically assess, from a social scientific perspective, the potential for harm in algorithmic design and machine learning systems. These questions have opened up discussions about how different mechanisms might help to pre-empt and address biases, including gendered ones, on social media platforms.

One approach is to conduct algorithmic impact assessments, which aim to identify potential bias and harm in algorithmic systems. Algorithmic impact assessments take different forms, including those that focus on eliminating gender-based prejudice and gendered harm. The Gender Shades Algorithm Audit, for example, assessed the performance of commercial facial recognition APIs in classifying faces by binary gender.[25] Other algorithmic impact assessments look at compliance with wider norms. This approach contextualises algorithms in organisational and individual behaviour. Given that algorithmic systems can be experienced in many different ways, as discussed in Chapter 3, attention to context appears key to understanding why and when certain biases might emerge. Algorithmic impact assessments also do not necessarily need to be focused on norms or bias, but can look more broadly at risks prior to operationalisation, or at impacts afterwards.[26]

While they are still relatively new, some governments have started to explore the potential of algorithmic impact assessments. There are draft regulations or legislation underway in Canada, New Zealand, the EU and the US (Moss et al., 2021a). Impact assessments are intended to unpack what the system does, and who should be responsible for remedying problems (Moss et al., 2021b).

Canada was an earlier adopter of algorithmic impact assessments. In April 2019, a Treasury Board Directive came into effect that outlined requires for algorithmic impact assessments for automated decision-making. The Canadian Algorithmic Impact Assessment Tool is a mandatory risk assessment tool for government agencies and for vendors serving government agencies (Government of Canada, n.d.).[27] A Data & Society report (Moss et al., 2021b; see also Moss et al., 2021a) has raised concerns about its unequal application; for example, the Canadian Department of Defence chose not to submit an assessment for use of algorithms to inform diversity in hiring (see also Data & Society, 2021).

---

25 Accuracy discrepancies were reduced a year after this was discovered.

26 The Ada Lovelace Institute and DataKind UK (2020) identify four categories of algorithmic impact assessments: (1) those focused on specific areas of bias; (2) those addressing wider compliance with regulation or norms (broader than bias audits); (3) risk assessments conducted prior to an algorithm's operation; and (4) impact evaluations after the algorithm is operationalised.

27 This is an online questionnaire, which results in an impact level and link to requirements under the Treasury Board's Directive on Automated Decision-Making. This came into force in April 2020.

Elsewhere, draft regulation by the EU would require developers of high-risk algorithmic systems to conform to performance standards (European Commission, 2021). In the US, draft legislation introduced to Congress in April 2019 would require commercial firms to conduct assessments for high-risk systems, including those using automated decision-making (United States Congress, 2019). It would also enable federal regulatory agencies to require impact assessments in areas where there is risk of discrimination, for instance in financial lending or real estate.

When considering the impact these approaches to social media regulation might have on gender norms, some important questions must be asked. Whose views are being heard and taken into account in these debates? What is the scope of harms being considered? Who is, and who is not, being empowered? For example, US and Canadian laws lack accountability forums that could improve their legitimacy, while the EU's approach relies largely on self-governance by developers, with few opportunities for public consultation over the values and aims of algorithmic systems (Moss et al., 2021b). The Data & Society report (ibid.) suggests that power dynamics in designing and implementing algorithmic impact assessments will vary across algorithmic systems and contexts.

Even further, while algorithmic impact assessments show some promise in considering the context and design of algorithmic systems, questions remain as to whether algorithms should be used to infer gender at all. Algorithmic impact assessments do not necessarily question whether gender can and should be translated into data or algorithms. Chapter 3 raised questions about biases inherent in the very labelling of gender as data. It could be argued that this process will always flatten and fundamentally misrepresent gender, which is a fluid, multifaceted and dynamic category. Systems that, at their core, are designed to simplify people's identities into binary classifications will fail to account for more marginalised and diverse human expressions (Vincent, 2021).

There is little clarity about when and to what extent algorithmic impact assessments could identify gender-related impacts in practice and how their findings might be meaningfully enforced to reduce harms. There are isolated cases of algorithmic audits revealing biases and contributing to changes in algorithmic design. Algorithmic audits can reveal biases in algorithmic processing, and there are instances where some algorithmic features have been removed as a result. One area that has come under criticism is the use of algorithms to recognise human emotions. Despite evidence that emotions are expressed in diverse ways and are difficult to read from facial expressions, such algorithms are increasingly and widely used (Murgia, 2021). After research and third-party algorithmic audits, one smaller firm, Hirevue, which uses software to assist in recruitment, removed visual analyses of job candidates (Zuloaga, 2021).

Yet, the extent to which algorithmic impact assessments could form the basis of more fundamental and widely implemented changes to the operation of algorithmic systems remains uncertain. More research is needed into how algorithmic impact assessments might influence the dynamic relationships (between algorithms, companies and platform users) that reproduce gender norms, and how these relationships might change over time, in different organisational contexts and in different jurisdictions.

## A human rights framework for regulation?

Advocates for gender equality have also used a human rights perspective to challenge uneven power dynamics and processes on social media platforms (Khan, 2021). The Human Rights Council stresses the importance of a human rights-based approach to internet provision and expansion (Khan, 2021: 56). In practice, this has been interpreted in different ways – for instance, through a focus on platforms as public spaces with public obligations, through putting human rights at the centre of platform decision-making and transparency, and through expanding user agency.

Colliver et al. (2021) challenge the notion of social media platforms as proprietary spaces, arguing instead that they are public spaces that should be easily navigable by a wide public. They argue that platforms' lack of transparency (about regulation, opportunities for complaint and redress, and the interrogation of algorithmically generated biases) in itself causes harm to users. Gillespie (2018) similarly suggests moving away from a focus on specific harms to a focus on platforms' obligations to the public. These approaches suggest that gendered harms on social media platforms should be addressed as part of a wider effort towards greater transparency in general, redistributing power in platform moderation.

David Kaye (former UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression) argues that company content moderation and its regulation needs to be founded in a commitment to apply international human rights law. This includes a form of public accountability where all rules, decisions and appeals about content are more transparent and governed by human rights. For instance, when a restriction is placed on the right to freedom of expression, Kaye argues, the reasons why this is necessary and proportionate must be clearly articulated (Kaye, 2019).

Spišák et al. (2021) also call for a human rights framework for regulation, including giving users the option to consent to accessing content, as opposed to operating within structures that erase content. These approaches, again, shift the focus away from instances of harm towards the process of developing regulation and its underlying framework and values. This suggests a promising approach to instituting processes that are oriented towards the needs of diverse users and what they envision for social media platforms. It also creates space to interrogate and address gender biases tied to the wider context in which algorithms are designed and operate (see also Sub-section 3.4 for a discussion of the protection of negative versus positive freedoms in platform moderation).

Another set of approaches focuses on expanding user agency on platforms, rather than removing content or restricting activity. Some advocates have recommended changes to technical processes that purportedly give users more time to recognise biases in content (such as gendered bias) and to make decisions about how to engage with that content. For example, this could include platforms building in more time for reflection, slowing down rather than speeding up information-sharing among users by adding more steps in user activity. This is termed 'friction' and its purpose is to encourage users to take more time to think before acting on social media (Khan, 2021). This could alter what users are likely to share, including, for example, content that is hurtful or biased along gender lines. Instagram and Twitter, for example, have introduced features aimed to encourage users to think before posting abusive content. Still, the effects of these processes on gender norms are yet to be studied. Additionally, without transparency, unidentified or subconscious bias – by companies, users or algorithms – might go untouched.

Others aim to improve user agency through the training and education of users. Here, there has been a specific focus on empowering women to navigate online communications safely and effectively. Media Pool, a Finnish organisation comprised of media officials, editors and other stakeholders, aims to provide support and redress to victims of state-sponsored harassment, including journalists and activists (Mediapooli, n.d.). It provides publications and courses to help media professionals technically prepare for crises and learn how to counter disinformation and online threats. Other initiatives focus on sensitising journalists to unintentional sexist and racist bias (Di Meco and Wilfore, 2021).

These suggestions do not necessarily point to alternatives to regulation, but they do suggest that external efforts to reshape relations on social media must take into account multiple factors, including power dynamics within the processes and the aims of regulation itself. Importantly, they indicate that human rights can be a fruitful reference point for new ways of approaching the regulation of social media platforms. A human rights-based approach expands the entry points for intervention; it can shift who is involved in the process of developing regulation and envisioning online environments that promote gender rights. A more systematic and critical review is needed to explore the potential of different human rights approaches and how to integrate them into platform operations and regulations.

## Regulating encrypted platforms and concerns over transparency

The previous sub-section identified the importance of platform transparency as part of incorporating human rights into platform regulation. However, transparency can be disempowering for users from marginalised groups or in authoritarian contexts, who rely on end-to-end encrypted communication channels to evade content monitoring and surveillance. These platforms, where only the sender and recipient can read the content, have become increasingly popular. WhatsApp (owned by Meta) has two billion monthly users (Statista, 2021c). In the late 2010s and early 2020s, Signal and Telegram attracted a surge of users seeking an alternative secure and encrypted platform. WeChat, based in China, is also popular, with 1.26 billion users in 2020 (Statista, 2021d). However, it is encrypted only from client to server, which means that its parent company, Tencent, can access all data (Associated Press, 2021).

Encryption and privacy protocols can benefit both the perpetrators of online harm and the victims. On the one side, encryption makes it more difficult to identify perpetrators, and on the other, it can provide a safe space for victims and survivors of violence (Aziz, 2017).

In the US, external concerns over content on encrypted platforms materialised around the 6 January 2021 attack on the US Capitol. After the attack, Telegram gained 25 million users (Gursky and Woolley, 2021).[28] In the US, content on encrypted platforms has fallen under the purview of external regulation in relation to sexually explicit content – for example, the US bipartisan EARN IT Act, which targets child sexual exploitation online.

In some instances, platforms and governments have collided over whether encrypted channels should be subject to external surveillance. In May 2021, WhatsApp sued the Indian government over internet laws that gave it scope to monitor content on encrypted channels. The government justified these laws

**28** The number of active Telegram users rose at this time to more than 500 million.

by claiming that encrypted channels were being used to incite mass violence (Ellis-Petersen, 2021). Around the same time, the Indian government also contested WhatsApp's updated privacy policy, which drew public attention to limited sharing of user account data with Facebook and its group of firms (Phartiyal, 2021). The debates over WhatsApp in India indicate the potential for regulations to cause harm. For example, marginalised groups' safety can be put at risk. However, who should decide when regulation is justified or when it might cause unjustified harm? This is a difficult question, especially as platforms operate in a variety of democratic and authoritarian contexts. Debates about regulating encrypted channels have tended to pitch privacy against harmful content. However, this itself might be unnecessarily restrictive. A 2021 report by Brookings argues that encryption can be protected while still dealing with problematic content by using technical interventions to limit its impact. For instance, platforms could focus on restricting specific actions, such as the forwarding of content (as WhatsApp has done) or bot-moderated activity (Gursky and Woolley, 2021).

The complex questions that arise in debates about the regulation of encrypted platforms indicate the need for caution when assessing the relationship between platform regulation and values such as gender equality (Khan, 2021). This is another area where more research is needed.

© Trismegist san/Shutterstock

## 4.3 Reflections

Given the global reach of social media platforms, across diverse audiences, contexts and jurisdictions, it is unlikely that one approach to regulation can resolve all harms. Users engage in diverse ways with one another and with the platforms, from within different sets of experiences (Gillespie, 2018). Platforms are constantly open to change, through technological innovation and in response to user activity.

Still, this overview of different approaches to the external regulation of social media platforms indicates some patterns that need to be taken into account when considering how to orient regulation towards gender equality. First, regulation faces a tension between the global reach of platforms and the centrality of the US legal system, and sometimes EU legislation, in shaping regulatory norms. This is tied to where most companies are headquartered. The dominance of US law, for example in safe harbour provisions, has framed the conceptualisation of social media platforms and what approaches might be possible globally. Second, while evidence is thin, there are indications that regulation can have unequal impacts on historically marginalised groups, including on women, LGBTQI+ and gender non-conforming people. The regulation of sexually explicit content is an area of concern, and LGBTQI+ people might be particularly affected in some jurisdictions.

External regulations that attend to the operation of technical processes (such as algorithmic impact assessments) or that are motivated by human rights suggest alternative approaches and possibilities. Could starting from concerns for human rights or technical processes, rather than content, alter the scope of user experiences online?

Approaches to regulation that focus on processes and human rights may be better equipped to confront the complex dynamics through which dominant gender norms are reproduced on social media. In order to reveal biases embedded in the design and use of platforms, the context in which they operate must also be taken into account. This is potentially a resource-intensive and locally oriented endeavour, but it is more likely to lead to forms of regulation that account for the dynamic and heterogeneous ways that social media infrastructure contributes to gender norms.

These approaches are complex, and more evidence is needed to understand how regulation intersects with the gendered dynamics between users and platforms explored in this report. This points to the importance of a gender lens, not only for understanding what happens on social media platforms, as explored in Chapter 3, but also for understanding the intended and unintended – and varied – consequences that regulations can have on the reproduction of gender norms.

" Could starting from concerns for human rights or technical processes, rather than content, alter the scope of user experiences online?

# 5 Conclusions and recommendations

The reach and use of social media platforms continues to expand, making it increasingly important to make sense of the constraints and opportunities embedded in platform infrastructure. The prevalence of hateful content toward women and non-conforming genders online raises concerns about whether and how social media is designed to promote gender equality and inclusion (Di Meco and Wilfore, 2021; Khan, 2021). What users do on each platform cannot fully explain how gender norms are reproduced in these spaces. Rather, gender norms on social media are shaped through dynamic and interdependent relationships between the platforms' business activities, their technological design, user activities, and even (although less examined) external regulation. This report contributes to a fuller understanding of the economic, technical and organisational infrastructure of different social media platforms and how these affect gender norms online. It also highlights opportunities and constraints for challenging existing inequalities and injustices.

Chapter 2 showed how specific features of social media platforms' business models, technology and organisational aspects shape what is likely to appear and be promoted to social media users. Chapter 3 explored how these features reproduce gender norms as they interact with user activity and experiences. In some instances, this results in the dominance of specific gender norms and the silencing of others. At the same time, marginalised groups have also been able to use platform features to strengthen their own agency online and express themselves in ways that might be constrained in other, offline forums. Chapter 4 considered how external regulation has a potentially complicated relationship with gender norms. It can both contribute to and disrupt gender-related harms and biases that emerge as a result of user–platform interactions. The subject of external regulation introduces additional questions about the legitimacy and inclusion of marginalised groups, both in how regulation is designed and in how it intersects with platform operations and use. These insights point to the need for greater attention to gender perspectives in research on the regulation of social media platforms.

In exploring the existing evidence on how social media platform infrastructure shapes gender norms, this report reveals some important areas for further research. The speed with which social media is expanding makes answering these questions urgent. Social media platforms are not merely sites where gender is performed or presented: they are themselves shifting gender norms online through the classification, analysis and promotion of content. This makes it important to develop a more complete understanding of how social media both constrains and creates possibilities for supporting gender equality online. There are any number of specific topics that could be explored in developing this understanding (many of which have been suggested in the report), from focusing on user experiences in specific geographical or linguistic contexts, to digging deeper into the organisational cultures of social media companies and exploring how internal company decisions are made.

To conclude, a successful research agenda will rest on three pillars: a transdisciplinary approach, a forward-looking analysis of regulation, and an intersectional global research and evidence base, explored more fully in the following list.

1. **Transdisciplinary approach:** At its foundation, any research programme on gender norms on social media will require a transdisciplinary approach that recognises the complex, reflexive and dynamic relationships between business processes, technological development, user behaviour and different forms of regulation. Social media platform infrastructure contains both stable and dynamic features – for example, algorithmic learning versus organisational structures. Studying gender norms online through a siloed disciplinary lens risks narrow and inaccurate conclusions that overemphasise certain features of platforms and how they operate. Understanding how shadowbanning takes place and its subsequent effects, for example, or the impacts of human and automated forms of content moderation, or the different ways that gender is performed on Tumblr compared to other platforms, requires teams of researchers who can bring together a range of specialised bodies of knowledge to create novel analytical frameworks for this new and highly dynamic space.

2. **Forward-looking analysis of regulation:** There is a need to understand the possibilities and implications of different approaches to regulating social media (again, using a transdisciplinary approach). Two linked research topics are suggested in the report. First, it would be valuable to develop techniques to assess the impact of platform models and behaviour. This could include: examining when and how algorithmic impact assessments could help to identify online harms and what they might miss; who should be responsible for harmful content and how they could be held accountable for remediation activities; and what alternative assessment methods exist. The second, broader, topic is to examine – in detail and over a long term – different approaches to regulating social media and their effectiveness in supporting gender equality online. National and regional (e.g. EU) jurisdictions regulate platform activity, content and liability differently. The ongoing evolution of social media platforms, from the creation of alt-right social media platforms to Facebook's aspirations of a metaverse that builds social connections into a virtual reality world, makes the regulatory space highly dynamic. This dynamism indicates both the challenges and the urgency of developing regulation. Research is required into distinctly inclusive and forward-looking regulatory approaches. Understanding how a human rights-centred framework for regulation could operate would be an appropriate place to start.

3. **An intersectional global research and evidence base:** Crucially, research must build a global evidence base and use an intersectional lens. The report highlights the enormous diversity of ways in which users experience and interact with platform infrastructure. However, studies of social media tend to look from a limited number of specific contexts – mainly in the Global North and in the English language – and presuppose universal 'features of communication with little cultural variation' (Pohjonen and Udupa, 2017: 1174; see also Gagliardone et al., 2021). At one level, this results in an incomplete picture of platforms and how their operations vary depending on combinations of algorithmic learning, user activity, regulation and socio-political contexts. At another level, and importantly in the context of this report, this fails to recognise the diversity of ways in which gender is expressed on platforms in different cultural contexts, and where and how women and people with non-conforming gender identities can utilise platform infrastructure to openly and safely engage online.

Finally, unpacking the relationship between social media platforms and gender norms, this report has sought to bridge technical and social perspectives of social media platforms, to enable an interdisciplinary dialogue about how social media platforms operate and why this matters. Hopefully, this will support greater dialogue among diverse researchers, software designers, policy-makers, activists and civil society that continues to unpack the decisions, relations and processes that feed into platform operations. Making social media into a place that is supportive of gender equality norms requires looking beyond what takes place online, to how and why different possibilities for use emerge.

# References


Abidin, C. (2019) 'Yes homo: gay influencers, homonormativity, and queerbaiting on YouTube' *Continuum* 33(5): 614–629 (https://doi.org/10.1080/10304312.2019.1644806).

Ada Lovelace Institute and DataKind UK (2020) *Examining the black box: tools for assessing algorithmic systems*. Ada Lovelace Institute and DataKind UK Report. London: Ada Lovelace Institute (www.adalovelaceinstitute.org/report/examining-the-black-box-tools-for-assessing-algorithmic-systems/).

Alaimo, C. and Kallinikos, J. (2017) 'Computing the everyday: social media as data platforms' *The Information Society* 33(4): 175–191 (https://doi.org/10.1080/01972243.2017.1318327).

ALIGN – Advancing Learning and Innovation on Gender Norms (n.d.) 'About norms: what are gender norms and how do they relate to social norms'. Webpage. ALIGN (www.alignplatform.org/about-norms).

Andrews, G. (2021) 'YouTube queer communities as heterotopias: space, identity and "realness" in queer South African vlogs' *Journal of African Cultural Studies* 33(1): 84–100 (https://doi.org/10.1080/13696815.2020.1792275).

Angwin, J. and Grassegger, H. (2017) 'Facebook's secret censorship rules protect white men from hate speech but not black children'. Propublica, 28 June (www.propublica.org/article/facebook-hate-speech-censorship-internal-documents-algorithms).

Apprich, C., Chun, W.H.K., Cramer, F. et al. (2018) *Pattern discrimination*. Minneapolis and London: University of Minnesota Press and Meson Press.

Are, C. (2020) 'How Instagram's algorithm is censoring women and vulnerable users but helping online abusers' *Feminist Media Studies* 20(5): 741–744 (https://doi.org/10.1080/14680777.2020.1783805).

Are, C. (2021) 'The shadowban cycle: an autoethnography of pole dancing, nudity and censorship on Instagram' *Feminist Media Studies* (https://doi.org/10.1080/14680777.2021.1928259).

Arvidsson A. (2016) 'Facebook and finance: on the social logic of the derivative' *Theory, Culture & Society* 33(6): 3-23 (https://doi.org/10.1177/0263276416658104).

Associated Press (2021) 'Signal: China appears to have blocked encrypted messaging app'. The Guardian, 16 March (https://www.theguardian.com/world/2021/mar/16/signal-blocked-china-encrypted-messaging-app).

Aziz, A.A. (2017) *Due diligence and accountability for online violence against women*. APC Issue Paper. Melville: Association for Progressive Communications (www.apc.org/sites/default/files/DueDiligenceAndAccountabilityForOnlineVAW.pdf).

Banet-Weiser, S. (2021a) 'Gender, social media, and the labor of authenticity' *American Quarterly* 73(1): 141–144 (https://doi.org/10.1353/aq.2021.0008).

Banet-Weiser, S. (2021b) 'Misogyny and the politics of misinformation' in H. Tumbler and S. Waisbord (eds) *The Routledge companion to media disinformation and populism*. London: Routledge: pp. 211–220.

Bateman, J., Thompson, N. and Smith, V. (2021) 'How social media platforms' community standards address influence operations'. Webpage/Report. Carnegie Endowment for International Peace, 1 April (https://carnegieendowment.org/2021/04/01/how-social-media-platforms-community-standards-address-influence-operations-pub-84201).

BBC News (2018) 'Facebook admits it was used to "incite offline violence" in Myanmar'. BBC News, 6 November (https://www.bbc.co.uk/news/world-asia-46105934).

Benjamin, R. (2019) *Race after technology: abolitionist tools for the new Jim Code*. Cambridge: Polity Press.

Bertaglia, T., Goanta, C. and Spanakis, G. (2020) 'Business model prevalence in influencer marketing on Instagram'. Presentation at the Workshop on Technology and Consumer Protection, 21 May, virtual conference (www.ieee-security.org/TC/SPW2020/ConPro/papers/bertaglia-conpro20-talk.pdf).

Bishop, S. (2018) 'Anxiety, panic and self-optimization: inequalities and the YouTube algorithm' *Convergence* 24(1): 69–84 (https://doi.org/10.1177/1354856517736978).

Bishop, S. (2021) 'Influencer management tools: algorithmic cultures, brand safety, and bias' *Social Media + Society* 7(1) (https://doi.org/10.1177/20563051211003066).

Bivens, R. (2017) 'The gender binary will not be deprogrammed: ten years of coding gender on Facebook' *New Media & Society* 19(6): 880–898 (https://doi.org/10.1177/1461444815621527).

Bivens, R. and Haimson, O.L. (2016) 'Baking gender into social media design: how platforms shape categories for users and advertisers' *Social Media + Society* 2(4) (https://doi.org/10.1177%2F2056305116672486).

Bodle, R. (2011) 'Regimes of sharing: open APIs, interoperability, and Facebook' *Information, Communication & Society* 14(3): 320–337 (https://doi.org/10.1080/1369118X.2010.542825).

Bridges, L.E. (2021) 'Digital failure: unbecoming the "good" data subject through entropic, fugitive, and queer data' *Big Data & Society* 8(1) (https://doi.org/10.1177%2F2053951720977882).

Bruns, A. (2019) 'After the "APIcalypse": social media platforms and their fight against critical scholarly research' *Information, Communication & Society* 22(11): 1544–1566 (https://doi.org/10.1080/1369118X.2019.1637447).

Bryan, A. (2019) 'Kuchu activism, queer sex-work and "lavender marriages," in Uganda's virtual LGBT safe(r) spaces' *Journal of Eastern African Studies* 13(1): 90–105 (https://doi.org/10.1080/17531055.2018.1547258).

Bucher, T. (2012) 'The friendship assemblage: investigating programmed sociality on Facebook' *Television & New Media* 14(6): 479–493 (https://doi.org/10.1177%2F1527476412452800).

Bumble (n.d.) 'We're launching Bumble Fund to invest in women founders'. Webpage. Bumble (https://bumble.com/en/the-buzz/bumble-fund).

Burgess, J., Cassidy, E., Duguay, S. et al. (2016) 'Making digital cultures of gender and sexuality with social media' *Social Media + Society* 2(4) (https://doi.org/10.1177%2F2056305116672487).

Burns, K.S. (2020) 'The history of social media influencers' in B. Watkins (ed.) *Research perspectives on social media influencers and brand communication*. Lanham: Lexington Books, pp. 1–21.

Bursztynsky, J. (2020) 'Facebook names two women to its board, nearing gender parity'. CNBC, 9 March (www.cnbc.com/2020/03/09/facebook-names-two-women-to-its-board-nearing-gender-parity.html).

Cadwalladr, C. and Graham-Harrison, E. (2018) 'Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach'. The Observer, 17 March (www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election).

Carr, C.T. and Hayes, R.A. (2015) 'Social media: defining, developing, and divining' *Atlantic Journal of Communication* 23(1): 46–65 (https://doi.org/10.1080/15456870.2015.972282).

Cavalcante, A. (2019) 'Tumbling into queer utopias and vortexes: experiences of LGBTQ social media users on Tumblr' *Journal of Homosexuality* 66(12): 1715–1735 (https://doi.org/10.1080/00918369.2018.1511131).

Chen, S.X. and Kanai, A. (2021) 'Authenticity, uniqueness and talent: Gay male beauty influencers in post-queer, postfeminist Instagram beauty culture' *European Journal of Cultural Studies*, online version (https://doi.org/10.1177/1367549421988966).

Cheney-Lippold, J. (2011) 'A new algorithmic identity: soft biopolitics and the modulation of control' *Theory, Culture & Society* 28(6):164–181 (https://doi.org/10.1177%2F0263276411424420).

Cho, A. (2018) 'Default publicness: queer youth of color, social media, and being outed by the machine' *New Media & Society* 20(9): 3183–3200 (https://doi.org/10.1177%2F1461444817744784).

Chun, W.H.K. (2011) *Programmed visions: software and memory*. Cambridge, MA: MIT Press.

Colliver, C., Comerford, M., King, J. et al. (2021) *Digital Policy Lab '20 companion papers*. Working Group Report. London: Institute for Strategic Dialogue (www.isdglobal.org/isd-publications/digital-policy-lab-20-companion-papers/).

Constine, J. (2017) 'Facebook drops no-vote stock plan, Zuck will sell shares to fund philanthropy'. TechCrunch, 22 September (https://techcrunch.com/2017/09/22/facebook-sharing/).

Costa, E. (2018) 'Affordances-in-practice: an ethnographic critique of social media logic and context collapse' *New Media & Society* 20(10): 3641–3656 (https://doi.org/10.1177/1461444818756290).

Cotter, K. (2019) 'Playing the visibility game: how digital influencers and algorithms negotiate influence on Instagram' *New Media & Society* 21: 895–913 (https://doi.org/10.1177%2F1461444818815684).

Couldry, N. and Mejias, U.A. (2019) *The costs of connection: how data is colonizing human life and appropriating it for capitalism*. Stanford, CA: Stanford University Press.

Crawford, K. and Gillespie, T. (2016) 'What is a flag for? Social media reporting tools and the vocabulary of complaint' *New Media & Society* 18(3): 410–428 (https://doi.org/10.1177%2F1461444814543163).

Crawford, K. and Paglen, T. (2019) 'Excavating AI: the politics of training sets for machine learning' *Stages* 9 (www.biennial.com/journal/issue-9/excavating-ai-the-politics-of-images-in-machine-learning-training-sets).

Crenshaw, K. (1991) 'Mapping the margins: intersectionality, identity politics, and violence against women of color' *Stanford Law Review* 43(6): 1241–1299 (https://doi.org/10.2307/1229039).

Data & Society (2021) '[ Audio ] DB145'. Transcript. Data & Society, 20 July (https://datasociety.net/wp-content/uploads/2021/07/Transcript-DB145-1.pdf).

De Veirman, M., Cauberghe, V. and Hudders, L. (2017) 'Marketing through Instagram influencers: the impact of number of followers and product divergence on brand attitude' *International Journal of Advertising* 36(5): 798–828 (https://doi.org/10.1080/02650487.2017.1348035).

Dean, J. (2008) 'Communicative capitalism: circulation and the foreclosure of politics' in M. Boler (ed.) *Digital media and democracy: tactics in hard times*. Cambridge, MA: MIT Press, pp. 101–122.

Department for Digital, Culture, Media & Sport and Home Office (2019) *Online harms*. White paper. London: HM Government (www.gov.uk/government/consultations/online-harms-white-paper).

DeVito, M.A. (2017) 'From editors to algorithms' *Digital Journalism* 5(6): 753–773 (https://doi.org/10.1080/21670811.2016.1178592).

Di Meco, L. and Wilfore, K. (2021) 'Canadian women leaders' digital defence initiative'. White Paper. Montreal: Montreal Institute for Genocide and Human Rights Studies, 3 June (https://issuu.com/migsinstitute/docs/whitepaper_final_version.docx).

D'Ignazio, C. and Klein, L.F. (2020) *Data feminism*. Cambridge, MA: MIT Press.

Dragiewicz, M., Burgess, J., Matamoros-Fernández, A. et al. (2018) 'Technology facilitated coercive control: domestic violence and the competing roles of digital media platforms' *Feminist Media Studies* 18(4): 609–625 (https://doi.org/10.1080/14680777.2018.1447341).

Drenten, J., Gurrieri, L. and Tyler, M. (2020) 'Sexualized labour in digital culture: Instagram influencers, porn chic and the monetization of attention' *Gender, Work & Organization* 27(1): 41–66 (https://doi.org/10.1111/gwao.12354).

Duffy, B.E. and Hund, E. (2015) '"Having it all" on social media: entrepreneurial femininity and self-branding among fashion bloggers' *Social Media + Society* 1(2) (https://doi.org/10.1177%2F2056305115604337).

Duffy, B.E. and Pruchniewska, U. (2017) 'Gender and self-enterprise in the social media age: a digital double bind' *Information, Communication & Society* 20(6): 843–859 (https://doi.org/10.1080/1369118X.2017.1291703).

Duguay, S. (2016) '"He has a way gayer Facebook than I do": investigating sexual identity disclosure and context collapse on a social networking site' *New Media & Society* 18(6): 891–907 (https://doi.org/10.1177%2F1461444814549930).

Ellis-Petersen, H. (2021) 'WhatsApp sues Indian government over 'mass surveillance' internet laws.' The Guardian, 26 May (www.theguardian.com/world/2021/may/26/whatsapp-sues-indian-government-over-mass-surveillance-internet-laws).

Eubanks, V. (2018) *Automating inequality: how high-tech tools profile, police, and punish the poor*. New York: St Martin's Press.

European Commission (2021) Proposal for a regulation of the European Parliament and of the council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts. Document 52021PC0206, 21 April (https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206).

Facebook (2021) *Annual diversity report*. Facebook Report (https://about.fb.com/wp-content/uploads/2021/07/Facebook-Annual-Diversity-Report-July-2021.pdf).

Federici, S. (2004) *Caliban and the witch: women, the body and primitive accumulation*. New York: Autonomedia.

Fink, M. and Miller, Q. (2014) 'Trans media moments: Tumblr, 2011–2013' *Television & New Media* 15(7): 611–626 (https://doi.org/10.1177/1527476413505002).

Fosch-Villaronga, E., Poulsen, A., Søraa, R.A. et al. (2021) 'A little bird told me your gender: gender inferences in social media' *Information Processing & Management* 58(3): 102541 (https://doi.org/10.1016/j.ipm.2021.102541).

Gagliardone, I., Diepeveen, S., Findlay, K. et al. (2021) 'Demystifying the COVID-19 infodemic: conspiracies, context, and the agency of users' *Social Media + Society* July: 1–16 (https://doi.org/10.1177/20563051211044233).

Gerrard, Y. and Thornham, H. (2020) 'Content moderation: social media's sexist assemblages' *New Media & Society* 22(7): 1266–1286 (https://doi.org/10.1177%2F1461444820912540).

Gibbs, S. (2017) 'Facebook bans women for posting 'men are scum' after harassment scandals'. The Guardian, 5 December (www.theguardian.com/technology/2017/dec/05/facebook-bans-women-posting-men-are-scum-harassment-scandals-comedian-marcia-belsky-abuse).

Gieseking, J.J. (2017) 'Messing with the attractiveness algorithm: a response to queering code/space' *Gender, Place & Culture* 24(11): 1659–1665 (https://doi.org/10.1080/0966369X.2017.1379955).

Gilbert, S. (2021) *Good data: an optimist's guide to our digital future*. London: Welbeck Publishing Group.

Gillespie, T. (2017) 'Governance of and by platforms' in J.E. Burgess, A.E. Marwick and T. Poell (eds) *The Sage handbook of social media*. London: Sage, pp. 254–278.

Gillespie, T. (2018) *Custodians of the internet: platforms, content moderation, and the hidden decisions that shape social media*. New Haven, CT: Yale University Press.

Gilroy, P. (1993) *The black Atlantic: modernity and double consciousness*. London: Verso.

Gorwa, R., Binns, R. and Katzenbach, C. (2020) 'Algorithmic content moderation: technical and political challenges in the automation of platform governance' *Big Data & Society* 7(1) (https://doi.org/10.1177%2F2053951719897945).

Government of Canada (n.d.) 'Algorithmic impact assessment tool'. Webpage. Government of Canada (www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html).

Green, A.I. (2013) 'Erotic capital' and the power of desirability: Why 'honey money' is a bad collective strategy for remedying gender inequality' *Sexualities* 16 (1-2): 137-158 (https://doi.org/10.1177/1363460712471109).

Gregg, M. and Andrijasevic, R. (2019) 'Virtually absent: The gendered histories and economies of digital labour' *Feminist Review* 123(1): 1–7 (https://doi.org/10.1177/0141778919878929).

Guo, L. and Johnson, B.G. (2020) 'Third-person effect and hate speech censorship on Facebook' *Social Media + Society* 6(2) (https://doi.org/10.1177%2F2056305120923003).

Gursky, J. and Woolley, S. (2021) *Countering disinformation and protecting democratic communication in encrypted messaging applications*. Brookings Briefing. Washington, DC: Brookings (www.brookings.edu/research/countering-disinformation-and-protecting-democratic-communication-on-encrypted-messaging-applications/).

Haimson, O.L., Dame-Griff, A., Capello, E. et al. (2021) 'Tumblr was a trans technology: the meaning, importance, history, and future of trans technologies' *Feminist Media Studies* 21(3): 345–361 (https://doi.org/10.1080/14680777.2019.1678505).

Haimson, O.L. and Hoffman, A.L. (2016) 'Constructing and enforcing "authentic" identity online: Facebook, real names, and non-normative identities' *First Monday*, 21 (6) (https://firstmonday.org/ojs/index.php/fm/article/download/6791/5521).

Hao, K. (2021) 'Facebook's ad algorithms are still excluding women from seeing jobs'. MIT Technology Review, 9 April (www.technologyreview.com/2021/04/09/1022217/facebook-ad-algorithm-sex-discrimination/).

Highfield, T. (2016) *Social media and everyday politics*. Cambridge: Polity Press.

Isaak, J. and Hanna, M.J. (2018) 'User data privacy: Facebook, Cambridge Analytica, and privacy protection' *Computer* 51(8): 56–59 (https://doi.org/10.1109/MC.2018.3191268).

Jankowicz, N., Hunchak, J., Pavliuc, A. et al. (2021) *Malign creativity: how gender, sex, and lies are weaponised against women online*. Wilson Center Report. Washington, DC: Wilson Center (www.wilsoncenter.org/publication/malign-creativity-how-gender-sex-and-lies-are-weaponized-against-women-online).

Kang, C. and Frenkel, S. (2018) 'Facebook says Cambridge Analytica harvested data of up to 87 million users'. New York Times, 4 April (www.nytimes.com/2018/04/04/technology/mark-zuckerberg-testify-congress.html).

Kaye, D. (2019) *Speech police: the global struggle to govern the internet*. New York: Columbia Global Reports.

Khan, I. (2021) *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*. United Nations Report A/76/258 (https://undocs.org/A/76/258).

Khattab, M. (2019) 'Synching and performing: body (re)-presentation in the short video app TikTok' *WiderScreen* 21(1–2) (http://widerscreen.fi/numerot/2019-1-2/synching-and-performing-body-re-presentation-in-the-short-video-app-tiktok/).

Kleeman, J. (2019) 'SNL producer and film-maker are latest to accuse YouTube of anti-LGBT bias'. The Guardian, 22 November (www.theguardian.com/technology/2019/nov/22/youtube-lgbt-content-lawsuit-discrimination-algorithm).

Klobuchar, A. (2021) 'Klobuchar, Kennedy, Manchin, Burr introduce bipartisan legislation to protect privacy of consumers' online data'. Press release, Amy Klobuchar/United States Senate, 20 May (www.klobuchar.senate.gov/public/index.cfm/2021/5/klobuchar-kennedy-manchin-burr-introduce-bipartisan-legislation-to-protect-privacy-of-consumers-online-data).

Klonick, K. (2021) 'Inside the making of Facebook's supreme court'. The New Yorker, 12 February (www.newyorker.com/tech/annals-of-technology/inside-the-making-of-facebooks-supreme-court).

Kurgan, L., Brawley, D., House, B. et al. (2019) 'Homophily: the urban history of an algorithm'. E-flux Architecture, October (www.e-flux.com/architecture/are-friends-electric/289193/homophily-the-urban-history-of-an-algorithm/).

Larkin, B. (2008) *Signal and noise: media, infrastructure, and urban culture in Nigeria*. London: Duke University Press.

Leerssen, P. (2015) 'Cut out by the middle man: the free speech implications of social network blocking and banning in the EU' *Journal of Intellectual Property, Information Technology and E-Commerce Law* 6(2): 99–119 (www.jipitec.eu/issues/jipitec-6-2-2015/4271/leerssen.pdf).

Ligaga, D. (2016) 'Presence, agency and popularity: Kenyan "socialites", femininities and digital media' *Eastern African Literary and Cultural Studies* 2(3–4): 111–123 (https://doi.org/10.1080/23277408.2016.1272184).

Lingel, J. and Golub, A. (2015) 'In face on Facebook: Brooklyn's drag community and sociotechnical practices of online communication' *Journal of Computer-Mediated Communication* 20(5): 536–553 (https://doi.org/10.1111/jcc4.12125).

LinkedIn (2020) 'Our 2020 workforce diversity report'. Press release. LinkedIn, 21 October (https://news.linkedin.com/2020/october/2020-workforce-diversity-report).

Lomborg, S. and Kapsch, P.H. (2020) 'Decoding algorithms' *Media, Culture & Society* 42(5): 745–761 (https://doi.org/10.1177/0163443719855301).

Lovelock, M. (2017) '"Is every YouTuber going to make a coming out video eventually?": YouTube celebrity video bloggers and lesbian and gay identity' *Celebrity Studies* 8(1): 87–103 (https://doi.org/10.1080/19392397.2016.1214608).

Madrigal, A.C. (2018) 'Inside Facebook's fast-growing content-moderation effort'. The Atlantic, 7 February (www.theatlantic.com/technology/archive/2018/02/what-facebook-told-insiders-about-how-it-moderates-posts/552632/).

Mahmood, L. (2021) 'Somali feminist: Facebook is being used to silence me.' BBC News, 8 September (www.bbc.co.uk/news/world-africa-58355603).

Mahoney, C. (2020) 'Is this what a feminist looks like? Curating the feminist self in the neoliberal visual economy of Instagram' *Feminist Media Studies* (https://doi.org/10.1080/14680777.2020.1810732).

Martinho, C. and Strickx, T. (2021) 'Understanding how Facebook disappeared from the internet'. The Cloudflare Blog, 4 October (https://blog.cloudflare.com/october-2021-facebook-outage/).

Marwick, A.E. and boyd, d. (2011) 'I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience' *New Media & Society* 13(1): 114–133 (https://doi.org/10.1177%2F1461444810365313).

McLelland, M. (2016) 'New media, censorship and gender: using obscenity law to restrict online self-expression in Japan and China' in L. Hjorth and O. Khoo (eds) *Routledge handbook of new media in Asia*. London: Routledge, pp. 130–141.

McNicol, A. (2013) 'None of your business? Analyzing the legitimacy and effects of gendering social spaces through system design' in M. Rasch and G. Lovink (eds) *Unlike us reader: social media monopolies and their alternatives*. Amsterdam: Institute of Network Cultures, pp. 200–219.

Mediapooli (n.d.) 'Mediapooli'. Website (www.mediapooli.fi/en/).

Meta (n.d.a) 'Facebook community standards'. Webpage. Meta (https://transparency.fb.com/en-gb/policies/community-standards/).

Meta (n.d.b) 'Government requests for user data'. Webpage. Meta (https://transparency.fb.com/data/government-data-requests/).

Meta (2020) 'Facebook UK gender pay gap report: April 2020'. Blog/Webpage. Meta (https://investor.fb.com/Facebook-UK-Gender-Pay-Gap-Report-April-2020/).

Meta (2021a) 'Facebook reports first quarter 2021 results'. Press Release. Meta, 28 April (https://investor.fb.com/investor-news/press-release-details/2021/Facebook-Reports-First-Quarter-2021-Results/default.aspx).

Meta (2021b) 'Our continuing commitment to transparency'. Blog. Meta, 19 May (https://about.fb.com/news/2021/05/transparency-report-h2-2020/).

Moran, R.E. (2020) 'Examining switching power: Mark Zuckerberg as a novel networked media mogul' *Information, Communication & Society* 23(4): 491–506 (https://doi.org/10.1080/1369118X.2018.1518473).

Morozov, E. (2019) 'Capitalism's new clothes'. The Baffler, 4 February (https://thebaffler.com/latest/capitalisms-new-clothes-morozov).

Moss, E., Singh, R., Watkins, E.A. et al. (2021a) 'Assembling accountability, from the ground up'. Data & Society: Points, 29 June (https://points.datasociety.net/assembling-accountability-from-the-ground-up-4655c492d0d0).

Moss, E., Watkins, E.A., Singh, R. et al. (2021b) *Assembling accountability: algorithmic impact assessment for the public interest*. Data & Society Report (https://datasociety.net/library/assembling-accountability-algorithmic-impact-assessment-for-the-public-interest/).

Murgia, M. (2021) 'Emotion recognition: can AI detect human feelings from a face?' Financial Times, 12 May (www.ft.com/content/c0b03d1d-f72f-48a8-b342-b4a926109452).

Murphy, H. (2021) 'Facebook warns of "significant" slowdown in revenue growth.' Financial Times, 28 July (www.ft.com/content/30f14409-438d-4066-a443-c09b3bceb5e6).

Nakamura, L. (2002) *Cybertypes: race, ethnicity, and identity on the internet*. New York: Routledge.

Noble, S.U. (2018) *Algorithms of oppression: how search engines reinforce racism*. New York University Press.

Nurik, C. (2019) '"Men are scum": self-regulation, hate speech, and gender-based censorship on Facebook' *International Journal of Communication* 13: 2878-2898 (https://ijoc.org/index.php/ijoc/article/view/9608/2697).

O'Meara, V. (2019) 'Weapons of the chic: Instagram influencer engagement pods as practices of resistance to Instagram platform labor' *Social Media + Society* 5(4) (https://doi.org/10.1177%2F2056305119879671).

O'Neil, C. (2016) *Weapons of math destruction: how big data increases inequality and threatens democracy*. New York: Penguin.

ONS – Office for National Statistics (2021) 'Gender pay gap in the UK: 2021'. Webpage. ONS (www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/bulletins/genderpaygapintheuk/2021).

Oversight Board (n.d.) 'Independent judgment. Transparency. Legitimacy.' Website. Oversight Board (https://oversightboard.com/).

Oversight Board (2020) 'Rulebook for case review and policy guidance', November (https://oversightboard.com/sr/rulebook-for-case-review-and-policy-guidance).

Parks, L. and Starosielski, N. (2015) *Signal traffic: critical studies of media infrastructures*. Champaign, IL: University of Illinois Press.

Petersson McIntyre, M. (2021) 'Commodifying feminism: economic choice and agency in the context of lifestyle influencers and gender consultants' *Gender, Work and Organization* 28(3): 1059–1078 (https://doi.org/10.1111/gwao.12627).

Phartiyal, S. (2021) 'India asks WhatsApp to withdraw its new privacy policy – sources'. Reuters, 20 May (www.reuters.com/technology/india-asks-whatsapp-withdraw-its-new-privacy-policy-sources-2021-05-20/).

Pilipets, E. and Paasonen, S. (2020) 'Nipples, memes and algorithmic failure: NSFW critique of Tumblr censorship' *New Media & Society* (https://doi.org/10.1177/1461444820979280).

Plantin, J.-C. and Punathambekar, A. (2019) 'Digital media infrastructures: pipes, platforms, and politics' *Media, Culture & Society* 41(2):163-174 (https://doi.org/10.1177%2F0163443718818376).

Pohjonen, M. and Udupa, S. (2017) 'Extreme speech online: an anthropological critique of hate speech debates' I*nternational Journal of Communication* 11: 1173-1191 (https://ijoc.org/index.php/ijoc/article/view/5843).

Posada, J. (2021) 'A new AI lexicon: labor'. AI Now Institute/Medium, 23 September (https://medium.com/a-new-ai-lexicon/a-new-ai-lexicon-labor-2db30b6b2349).

Privacy International (2018) 'Why and how GDPR applies to companies globally'. Blog. Privacy International, 25 May (https://privacyinternational.org/long-read/2207/why-and-how-gdpr-applies-companies-globally).

Quick, T. (2021) 'Influencer culture and algorithmic apartheid on gay Instagram' AoIR Selected Papers of Internet Research, virtual event, 13-16 October (https://doi.org/10.5210/spir.v2021i0.12019).

Reese, H. and Heath, N. (2016) 'Inside Amazon's clickworker platform: how half a million people are being paid pennies to train AI'. TechRepublic, 16 December (www.techrepublic.com/article/inside-amazons-clickworker-platform-how-half-a-million-people-are-training-ai-for-pennies-per-task/).

Rhee, L., Bayer, J.B., Lee, D.S. et al. (2021) 'Social by definition: how users define social platforms and why it matters' *Telematics and Informatics* 59: 101538 (https://doi.org/10.1016/j.tele.2020.101538).

Roberts, S.T. (2018) 'Digital detritus: "error" and the logic of opacity in social media content moderation' *First Monday* 23(3) (https://doi.org/10.5210/fm.v23i3.8283).

Romm, T. (2021) 'Amazon, Facebook, other tech giants spent roughly $65 million to lobby Washington last year'. Washington Post, 22 January (www.washingtonpost.com/technology/2021/01/22/amazon-facebook-google-lobbying-2020/).

Romo, V. (2021) 'Whistleblower's testimony has resurfaced Facebook's Instagram problem'. NPR, 5 October (www.npr.org/2021/10/05/1043194385/whistleblowers-testimony-facebook-instagram?t=1633982889234).

Rosenberg, E. (2019) 'A right wing YouTuber hurled racist, homophobic taunts at a gay reporter. The company did nothing'. Washington Post, 5 June (www.washingtonpost.com/technology/2019/06/05/right-wing-youtuber-hurled-racist-homophobic-taunts-gay-reporter-company-did-nothing/).

Salty (2019) 'Exclusive: an investigation into algorithmic bias in content policing on Instagram'. Salty, October (https://saltyworld.net/algorithmicbiasreport-2/).

Sances, M. (2021) 'Missing the target? Using surveys to validate social media ad targeting' *Political Science Research and Methods* 9(1): 215–222 (https://doi.org/10.1017/psrm.2018.68).

Siegel, R. (2019) 'Tumblr once sold for $1.1 billion. The owner of WordPress just bought the site for a fraction of that'. Washington Post, 13 August (www.washingtonpost.com/technology/2019/08/13/tumblr-once-sold-billion-owner-wordpress-just-bought-site-fraction-that/).

Simonite, T. (2021) 'Facebook is everywhere; its moderation is nowhere close'. Wired, 25 October (www.wired.com/story/facebooks-global-reach-exceeds-linguistic-grasp/).

Sonnemaker, T. (2020) 'Facebook just named two women to its board as it seeks gender parity — here are 13 tech companies that have recently diversified their boardrooms'. Insider, 16 March (www.businessinsider.com/13-tech-companies-added-women-corporate-board-diversity-2019-2020-3).

Southerton, C., Marshall, D., Aggleton, P. et al. (2021) 'Restricted modes: social media, content classification and LGBTQ sexual citizenship' *New Media & Society* 23(5): 920-938 (https://doi.org/10.1177/1461444820904362).

Spišák, S., Pirjatanniemi, E., Paalanen, T. et al. (2021) 'Social networking sites' gag order: commercial content moderation's adverse implications for fundamental sexual rights and wellbeing' *Social Media + Society* 7(2) (https://doi.org/10.1177/20563051211024962).

Statista (2021a) 'Distribution of Facebook employees worldwide from 2014 to 2020, by gender'. Webpage. Statista, 27 Jan (www.statista.com/statistics/311827/facebook-employee-gender-global/).

Statista (2021b) 'Distribution of influencers creating sponsored posts on Instagram worldwide in 2019, by gender'. Webpage. Statista, 13 August (www.statista.com/statistics/893749/share-influencers-creating-sponsored-posts-by-gender/).

Statista (2021c) 'Number of monthly active WhatsApp users worldwide from April 2013 to March 2020'. Webpage. Statista, 9 December (www.statista.com/statistics/260819/number-of-monthly-active-whatsapp-users/).

Statista (2021d) 'Number of monthly active WeChat users from 2nd quarter 2011 to 3rd quarter 2021'. Webpage. Statista, 6 December (www.statista.com/statistics/255778/number-of-active-wechat-messenger-accounts/).

Stokel-Walker, C. (2019) 'Facebook's ad data may put millions of gay people at risk'. New Scientist, 24 August (www.newscientist.com/article/2214309-facebooks-ad-data-may-put-millions-of-gay-people-at-risk/).

Stone, B. (2009) 'Facebook will form 2 classes of stock'. New York Times, 24 November (www.nytimes.com/2009/11/25/technology/internet/25facebook.html).

Strimpel, Z. (2021) 'Bumble's "feminism" is half-baked'. Spectator, 17 February (www.spectator.co.uk/article/bumble-s-feminism-is-half-baked).

Tamale, S. (2003) 'Out of the closet: unveiling sexuality discourses in Uganda' *Feminist Africa* 2: 42–49 (www.agi.ac.za/sites/default/files/image_tool/images/429/feminist_africa_journals/archive/02/fa_2_standpoint_3.pdf).

Taylor, J. (2021) 'Facebook outage: what went wrong and why did it take so long to fix after social platform went down?' The Guardian, 5 October (www.theguardian.com/technology/2021/oct/05/facebook-outage-what-went-wrong-and-why-did-it-take-so-long-to-fix).

Telford, T. (2021) 'Twitter algorithms amplify conservative content more than that of the political left, researchers find'. Washington Post, 22 October (www.washingtonpost.com/business/2021/10/22/twitter-algorithm-right-leaning/).

The Economist Intelligence Unit (2021) 'Measuring the prevalence of online violence against women' Infographic. The Economist, 1 March (https://onlineviolencewomen.eiu.com/).

The Good Robot (2021) 'Anita Williams on countering online sex abuse'. Podcast. The Good Robot/Listen Notes, 1 June (www.listennotes.com/podcasts/the-good-robot/anita-williams-on-countering-63gT9ZPmwWs/).

Tiidenberg, K. and van der Nagel, E. (2020) *Sex and social media*. Bingley: Emerald Publishing.

TikTok (2021) 'TikTok UK gender pay gap report (April 5th, 2020 snapshot)'. Blog/Webpage. TikTok, 1 April (https://newsroom.tiktok.com/en-gb/tiktok-uk-gender-pay-gap-report-2019-20-pay-period).

Twitter (n.d.) 'Inclusion, diversity, equity, and accessibility'. Webpage. Twitter (https://careers.twitter.com/en/diversity.html#Leadership).

United States Congress (2019). H.R.2231 - Algorithmic Accountability Act of 2019 (www.congress.gov/bill/116th-congress/house-bill/2231).

Vaiciukynaite, E. (2019) 'Men or women? Neuro-marketing study of social media influencers'. Presentation at the 6th European Conference on Social Media, 13–14 June 2019, University of Brighton, Brighton.

van der Vlist, F.N. and Helmond, A. (2021) 'How partners mediate platform power: mapping business and data partnerships in the social media ecosystem' *Big Data & Society* 8(1) (https://doi.org/10.1177%2F20539517211025061).

Van Dijck, J. (2020) 'Seeing the forest for the trees: visualizing platformization and its governance' *New Media & Society* 23(9): 2801–2819 (https://doi.org/10.1177/1461444820940293).

Van Driel, L. and Dumitrica, D. (2021) 'Selling brands while staying "authentic": the professionalization of Instagram influencers' *Convergence* 27(1): 66–84 (https://doi.org/10.1177%2F1354856520902136).

Vincent, J. (2021) 'Automatic gender recognition tech is dangerous, say campaigners: it's time to ban it'. The Verge, 14 April (www.theverge.com/2021/4/14/22381370/automatic-gender-recognition-sexual-orientation-facial-ai-analysis-ban-campaign).

Wakabayashi, D. (2020) 'Legal shield for social media is targeted by lawmakers'. New York Times, 28 May (www.nytimes.com/2020/05/28/business/section-230-internet-speech.html).

Walton, J. (2022) 'Twitter vs. Facebook vs. Instagram: What's the Difference?' Investopedia, 13 January (www.investopedia.com/articles/markets/100215/twitter-vs-facebook-vs-instagram-who-target-audience.asp).

Washington, K. and Marcus, R. (2022 – in press) *Can social media and online activism shift gender norms?* ALIGN Report. London: ODI (www.alignplatform.org/resources/report-can-social-media-and-online-activism-shift-gender-norms).

West, S.M. (2017) 'Raging against the machine: network gatekeeping and collective action on social media platforms' *Media and Communication* 5(3): 28–36 (www.cogitatiopress.com/mediaandcommunication/article/view/989/989).

West, S.M. (2020) 'Redistribution and rekognition: a feminist critique of algorithm fairness' *Catalyst: Feminism, Theory, Technoscience* 6(2): 1–24 (https://catalystjournal.org/index.php/catalyst/article/view/33043).

Williams, M. (2021) 'Facebook diversity update: increasing representation in our workforce and supporting minority-owned businesses'. Meta, 15 July (https://about.fb.com/news/2021/07/facebook-diversity-report-2021/).

Wilson, S., Kaur, K., Mogan, S. et al. (2018) 'Empowering women through Instagram posts: a case study analysis of #Sareesandstories'. Presentation at Kanita International Conference on Gender Studies, 27–28 November, Centre for Research on Women and Gender (KANITA), Universiti Sains Malaysia, Pulau Pinang (http://eprints.usm.my/49444/1/Pages%20from%20PROCEEDINGS-4%20%281%29-2%20wilson.pdf).

Young, L. (2019) 'How much do influencers charge?' Klear, 16 May (https://klear.com/blog/influencer-pricing-2019/).

Zuboff, S. (2019) *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. London: Profile books.

Zuloaga, L. (2021) 'Industry leadership: new audit results and decision on visual analysis'. HireVue, 12 January (www.hirevue.com/blog/hiring/industry-leadership-new-audit-results-and-decision-on-visual-analysis).

ALiGN

## About ALIGN

ALIGN is a digital platform and programme of work that is creating a global community of researchers and thought leaders, all committed to gender justice and equality. It provides new research, insights from practice, and grants for initiatives that increase our understanding of – and work to change – discriminatory gender norms. Through its vibrant and growing digital platform, and its events and activities, ALIGN aims to ensure that the best of available knowledge and resources have a growing impact on harmful gender norms.

## Disclaimer

This document is an output of Advancing Learning and Innovation on Gender Norms (ALIGN). The views expressed and information contained within are not necessarily those of or endorsed by ODI, Global Affairs Canada or the Ford Foundation, and accepts no responsibility for such views or information or for any reliance placed on them.


## ALIGN Programme Office

ODI
203 Blackfriars Road
London SE1 8NJ
United Kingdom
Email: **align@odi.org.uk**
Web: **www.alignplatform.org**




alignplatform.org


ALIGN is a research platform currently supported by the Government of Canada (through Global Affairs Canada) and is led by ODI.


ODI Think Change