CHAPTER 5

Shinigami Eyes and Social Media Labeling as a Technology for Self-care

Henrik Skaug Sætra and Jo Ese

CONTENTS

5.1 INTRODUCTION

Since its inception, the role of social media as part of the public sphere (Habermas, 1991) has been debated. Ranging from the panegyric praise of the emancipating role of Twitter and Facebook during the Arab Spring to the condemnation of the corrosive effects of the very same platforms during the last two US presidential campaigns, social media as part of our infrastructure for a sustainable and sound public debate has been contested (Kruse et al., 2018; Vaidhyanathan, 2018). Furthermore, social media platforms are constantly developing, with changing algorithms nudging the public debate in different directions, potentially profoundly influencing interactions, the public debate, and eventually how our democracies and societies are structured.

However, if there ever existed such a thing as a Habermasian public sphere in present-day society, the concept is certainly challenged by recent developments. In this chapter,

DOI: 10.1201/9781003325086-5

we examine one particular challenge exemplified by the *Shinigami Eyes* web browser extension (Kiran, 2022). The extension is developed as a tool to aid trans people and their allies in identifying and avoiding transphobic users and content, while easily recognizing trans-friendly users. However, the add-on has received criticism due to its labeling of social media users without the users knowing about or agreeing to the labeling, and some feel that such a tool undermines the very idea of a public sphere and free and open public debate.

Although society's treatment and acceptance of trans individuals have gone through major developments in the last decades, marginalization, exclusion, abuse, and harassment of trans individuals are still widespread (Ciszek et al., 2021). As with other forms of harassment, a lot of the anti-trans harassment is found online. Social media is a prominent part of the lives of modern individuals in major parts of the world, and LGBTQ+ people are also actively using social media to navigate their existence and construct their identities (Jenzen, 2017; Lucero, 2017; Southerton et al., 2021). However, in this chapter, we discuss how these individuals, and trans people in particular, experience harassment and hostility on social media (Jenzen, 2017), which when seen in conjunction with other forms of marginalization poses clear threats relevant to a number of aspects related to sustainable development and the Sustainable Development Goals (SDGs) (Ongsupankul, 2019).

For example, "full and effective participation and equal opportunities for leadership at all levels of decision-making in political, economic and public life" is a target of SDG 5 (United Nations, 2015), and while this might seem to offer opportunities to promote the rights of LGBTQ+ people, we show how these individuals are erased and omitted from major frameworks such as the SDGs. Nevertheless, we might interpret the SDGs more broadly (Dorey & O'Connor, 2016), and if so point out how, for example, participation in public life requires an online presence and online participation. If these arenas are perceived as threatening and dangerous, it can be argued that these threats undermine sustainable development. A tool like Shinigami Eyes, which potentially helps trans individuals avoid content perceived as harmful, thus makes it easier to participate in this important domain of public life (Lucero, 2017).

However, public life is arguably also about the exchange of arguments and ideas, including facing, experiencing, arguing against, and sometimes learning from ideas and opinions that we disagree with. Although a technology like Shinigami Eyes protects individuals from content they find uncomfortable, some fear it could also insulate individuals from opinions that do not resemble their own. Without communication and exchanges of ideas, arguments, and experiences between different groups, conflict and polarization may arise. Furthermore, those who are labeled also have certain rights and interests. For example, SDG 16.10 specifically addresses fundamental freedoms like freedom of thought and freedom of expression, which align with arguments used by the Norwegian data protection authority to *ban* Shinigami Eyes from processing data in Norway (Datatilsynet, 2022a). Furthermore, the labeling of individuals challenges the Habermasian ideal of the public sphere as a place where the *content* of a message – and not the identity of the speaker – is what should be considered, an ideal also promoted by Merton (1942) in the domain of science.

In this chapter, we examine the sustainability-related implications of trans people's online lives and Shinigami Eyes labeling, and raise the following questions:

1. How do issues related to the rights of transgender people relate to sustainable development and the SDGs?
2. Is social media potentially so hateful and violent toward vulnerable groups that the kind of labeling described in this chapter is necessary and legitimate?
3. Which conflicts between the rights of different individuals and different SDGs are central to understanding the controversy surrounding Shinigami Eyes?

5.2 SUSTAINABILITY, LGBTQ+ RIGHTS, AND SOCIAL MEDIA

Sustainability, as discussed in Chapter 2, encompasses economic, social, and environmental issues. The broader notion of social sustainability, including the principles of inclusion, safety, and justice for all, is clearly amenable to include issues related to, for example, discrimination and hate-speech based on sexual orientation, gender identity and expressions, and sexual characteristics (SOGIESC).[1] One might, then, be inclined to assume that the SDGs also provide protection for people experiencing exclusion and discrimination based on SOGIESC. Such an assumption would, however, be unwarranted.

Quite the contrary, the SDGs do not mention SOGIESC or elements of it anywhere (Mills, 2015; Ongsupankul, 2019; United Nations, 2015), as proponents of LGBTQ+ rights and equality have long pointed out (Dorey & O'Connor, 2016; Ongsupankul, 2019). Some go so far as to state that, for example, gender-diverse and trans people are "completely omitted and erased from the scope" of key SDGs, such as SDG 5 (Matthyse, 2020). Others support the notion that LGBTQ+ people are erased in Agenda 2030, and also that their sexual rights are, problematically, omitted from the SDGs (Logie, 2021). Politics is a crucial part of the SDGs and similar frameworks (Majeedullah et al., 2016), and while efforts to include all or parts of the LGBTQ+ community in international frameworks such as the SDGs achieve intermittent success, the advances made in negotiations are often defeated before final publication. Ongsupankul (2019), for example, relates how LGBT rights were said to be "off the table" in the lead-up to Agenda 2030, as the need to compromise with countries where LGBTQ+ people and their concerns are not accepted or recognized thoroughly complicates the negotiation process.

Despite the lack of explicit mention, many are arguing that SOGIESC is encompassed by the core goal of the SDGs, namely to *leave no one behind* (United Nations, 2015). This concept is consequently highlighted by many of those exploring the rights of LGBTQ+ people from a sustainability perspective (Logie, 2021; Ongsupankul, 2019). The most optimistic accounts of SOGIESC and the SDGs emphasize this and argue that while SOGIESC issues are not explicitly mentioned, LGBTQ+ people are included through what is referred to as an "other" status in various reports from, for example, the United Nations, the Human Rights Council, and the World Health Organization (Divan et al., 2016; Logie, 2021; Mills, 2015).

Stonewall International has released a guide for LGBT inclusion and the SDGs, and while they state that the SDGs "could have gone further", this particular publication highlights

how the SDGs can in fact be used to improve the situation of LGBTQ+ people through the focus on leaving no one behind and equality for *all* in Agenda 2030 (Dorey & O'Connor, 2016). Their analysis of the most relevant goals is summarized in the following.

SDG 1 aims to end poverty in all its forms everywhere and entails social protection systems for all (target 1.3), making sure that all have equal rights to economic resources, access to services, and control over land and property (target 1.4). As LGBTQ+ people are excluded and discriminated in a wide range of settings – in education and work – they are economically vulnerable, and there is a need for more research on economic discrimination, including SOGIESC considerations in development projects, and social assistance projects (Dorey & O'Connor, 2016).

SDG 3 is about the promotion of healthy lives and well-being for all, and *universal* access to sexual and reproductive healthcare services (target 3.7) and universal health coverage for all (target 3.8) are of particular relevance. LGBTQ+ people's access to good and non-discriminatory health services is key, and this entails education and prevention measures, proper help for trans people seeking to transition, and not least mental health services and support (Dorey & O'Connor, 2016).

Quality education for all is the aim of SDG 4, and LGBTQ+ people experience bullying and exclusion throughout the educational system, leading many to dropout and miss future opportunities (Majeedullah et al., 2016). Increased awareness and competence for teachers and counselors, active policies against homo-, bi-, and transphobic bullying, inclusion of LGBTQ+ inclusive and positive material in education, and general efforts to foster more inclusive cultures are among the key actions required for achieving this goal (Dorey & O'Connor, 2016).

SDG 5 is also mentioned, despite this goal explicitly and exclusively targeting *women* and *girls* (Ongsupankul, 2019). Dorey and O'Connor (2016) here emphasizes intersectionality and how lesbian, trans, and bi women are particularly vulnerable to discrimination and violence. In a critical analysis of SDGs and transgender equality, Matthyse (2020) argues that the omission of gender-diverse and trans people in SDG 5 has been achieved by the patriarchy to continue gendered oppression and inequality, and highlights the unfortunate impacts of such erasure in the global framework designed to include all. They proceed to argue that an "authentic version of gender equality to achieve freedom from oppression on the grounds of gender takes into account a multiplicity of gender identities and gender expressions" (Matthyse, 2020), something also supported by others (Majeedullah et al., 2016).

One of the goals often presented as the most obvious candidate for addressing SOGIESC challenges is SDG 10 (Majeedullah et al., 2016), which aims to reduce inequality within and among countries. Target 10.2 states that the goal is to empower and promote the inclusion of *all*, irrespective of a number of explicitly mentioned characteristics (age, race, sex, etc.) "or other status", which is taken to include LGBTQ+ people (Logie, 2021). Target 10.3 proceeds to state that equal opportunity should be promoted by eliminating discriminatory laws, policies, and practices. This goal is consequently crucial for addressing the discrimination of LGBTQ+ people through discriminatory laws and policies. Examples of such policies include requiring trans people to undergo sterilization in order to change gender, not providing opportunities for trans people to change legal gender, prohibiting same-sex

or other relations, prohibiting the establishment of LGBTQ+ groups and organizations, and excluding LGBTQ+ people from social and health services. Dorey and O'Connor (2016) argue that generating awareness of how LGBTQ+ people are included in "other" status is key to achieving progress in this domain.

SDG 11 is also relevant, as it relates to the inclusiveness and safety of cities and human settlements. This could arguably be related to SOGIESC-related violence and exclusion, but Dorey and O'Connor (2016) focus on target 11.1 and access to affordable housing and basic services, as many LGBTQ+ people are forced to leave their families and homes and experience high prevalence of homelessness. Violence is dealt with under SDG 16, which is another goal with high thematic overlap with sexuality- and gender-related challenges (Majeedullah et al., 2016). SDG 16 has targets related to reducing all forms of violence and death threats (target 16.1) and ensuring access to justice (16.3). While all forms of violence are abhorrent, Dorey and O'Connor (2016) highlight the challenges related to how states are not taking violence against LBGTQ+ people seriously, and that these individuals in many areas consequently cannot rely on the state's protection. This makes these individuals more vulnerable to human rights abuse, and taking action to increase awareness of and the prioritization of hate crimes is crucial for reaching SDG 16 (Dorey & O'Connor, 2016).

It is important to be aware of country-to-country and even community-to-community differences regarding the situation of LGBTQ+ people. Even if some readers consider the situation for these individuals in their society to be relatively safe and that they have equal rights and protections as others, we are here considering the global context and how people from many different countries are coming together on platforms and social media that are often not clearly regulated through national legislation. On such platforms, those who do *not* respect LGBTQ+ rights cross paths with others who might *have* rights nationally. This gives rise to the conflict of rights and interests discussed later. Those who do not have rights in their homeland, such as people in Tunisia where homosexuality is criminalized and any offense is retained in a person's criminal record for 5 years (Ongsupankul, 2019), may be able to find support and the means to partly fight back through social media as they can connect with like-minded people and share and develop tools and systems (Haimson et al., 2020; Haimson et al., 2021).

Despite various efforts and initiatives to improve their situation, the reality is that LGBTQ+ people are amongst the most vulnerable individuals in many modern societies. And amongst this group, trans people are the most vulnerable of all (Jenzen, 2017; Ongsupankul, 2019). Trans people experience "extreme social exclusion" which result in increased vulnerability to physical and mental health conditions, discrimination in education and work, and a "general loss of opportunities for economic and social advancement" (Divan et al., 2016).

To add to this trans people who experience resentment, prejudice, and threatening environments often find themselves in criminalized contexts, discouraging them from seeking the aid of police and barring them from other avenues of justice (Divan et al., 2016). Trans people are often criminalized and misunderstood, deprived of basic services (Matthyse, 2020), and, for example, trans women have alarmingly low life expectancies – as low as 35 years old in the Americas (Inter-American Commission on Human Rights, 2018).

Remedies for these vulnerabilities require action at numerous levels, including legal reform and the criminalization of various forms of violence and hate speech against trans people, getting rid of laws that criminalize and expose trans people, and "general advocacy to sensitize the ill-informed about trans issues and concerns" (Divan et al., 2016). This involves community-level reforms of the societal attitudes and social practices that engender the marginalization of trans people and others (Majeedullah et al., 2016).

Summing up, trans people live in "extremely hostile contexts" (Divan et al., 2016) and live with a higher risk of experiencing various forms of violence and hate crimes (Matthyse, 2020). That a framework such as Agenda 2030 and the SDGs fail to mention and explicitly emphasize their rights highlights the potential need for *other*, and potentially untraditional, means to improve and safeguard the position of these individuals, both by the individuals themselves and by others who support their cause. While institutions, local communities, and families are often highlighted as arenas in which gender and sexuality norms are policed (Majeedullah et al., 2016), public spheres and online communities are potentially equally important arenas both of moral policing and crucial potential arenas for escape from the aforementioned hostile arenas (Haimson et al., 2020; Jenzen, 2017). Furthermore, as many sexual minorities live in contexts in which they cannot invoke legal rights or prefer not to report crimes committed against them out of fear of being arrested (Ongsupankul, 2019), self-care of various kinds becomes necessary (Edmond, 2022). One way for the trans community to protect themselves online is the marking and filtering of perceived harmful individuals in social media, to which we now turn.

5.3 THE CASE: SHINIGAMI EYES AND SOCIAL MEDIA LABELING

We focus on one specific technological response to the preceding challenges LGBTQ+ people experience online, namely a browser add-on named *Shinigami Eyes* (Kiran, 2022). The add-on is free and distributed on GitHub, and it works for Chrome and Firefox browsers. What it does is highlight "transphobic and trans-friendly social network pages and users with different colors" (Kiran, 2022). The publisher is an anonymous self-declared trans person publishing under the nickname *Kiran*. They describe making this extension to alleviate some of the uncertainty trans people experience when they face communities with members whose interests, views, and opinions might be hostile to trans people. This is particularly difficult, Kiran writes, when dealing with groups of intersectional interests, such as the feminist, lesbian, and atheist communities (Kiran, 2022).

> The purpose of this extension is to make transgender people feel more confident towards people, groups, and pages they can trust, and to highlight possible interactions with the trans-hostile ones.
>
> (Kiran, 2022)

Intersectionality refers to how minority characteristics seldom come alone, and how, for example, race and gender interact in shaping the experiences of marginalized people (Crenshaw, 1990; Matthyse, 2020). SOGIESC, and being trans as highlighted by Kiran, provides additional minority characteristics that interact with, for example, gender. While

feminists and trans people share a broad experience of being marginalized, the interactions between trans people and sections of the feminist community are often quite hostile (Hines, 2019). Dealing with groups that are uniformly and openly hostile to trans people is described by Kiran as relatively easy, but dealing with groups with members who are seemingly progressive but at the same time hostile to trans people is difficult. This generates insecurity, as trans people are left guessing who their allies and enemies are (Kiran, 2022). In such a process, each individual is left to keep track of whom to trust and not which, needless to say, is a tall order on the supported platforms Facebook, Twitter, Reddit, Tumblr, Medium, YouTube, Wikipedia, and search engine results.

Edmond (2022) provides support for the usefulness of Shinigami Eyes in their account of how a user of the extension states that Shinigami Eyes does in fact aid in conducting the "vetting process" they were already undertaking (Edmond, 2022). As also emphasized by Haimson et al. (2021), tags and filtering "will be essential to any future trans technology", and Shinigami Eyes provides one way to achieve this on platforms not explicitly made for this group.

Taking one step back, the add-on has taken its name from the popular Japanese manga and anime show called *Death Note* (original Japanese title *Desu Nōto*) (Nakatani et al., 2006–2007; Ohba & Obata, 2003–2008). In this show, death gods – the *Shinigami* – enter the world of the humans and provide a select few with notebooks called Death Notes. These notebooks provide their human owners with the power to end the lives of others merely by writing their names in the notebook while imagining their faces. The Shinigami have special eyes that enable them to see the names and remaining life span of the humans they encounter. As the notebook requires someone's real name for them to die, this is an attractive ability for death note holders. The Shinigami consequently offer the human notebook owners a deal: They can have *Shinigami Eyes* in exchange for half of their remaining life span. This would then enable them to see the names and life span of their fellow humans in bright red and in real time. This brief background should suffice to show that the background of the add-on's name is quite sinister, and that its functionality is only superficially connected to the function of Shinigami Eyes in the anime. The publisher's nickname is also presumably derived from the series' protagonist's nickname – *Kira*.

The browser extension, when installed, colors people's usernames either red or green depending on whether they have been labeled trans-hostile or trans-friendly. Figure 5.1 shows how this appears on the authors' Twitter, where the usernames and other information are clearly colored and consequently stand out from other users.

The initial extension marked people based on the author's manual labeling mixed with "machine learning", and people can now contribute their own labels as well (Kiran, 2022). Figure 5.2 shows the user interface that allows for the labeling of users, over a post where the labeling of Margaret Atwood is being discussed.[2]

The main input for labeling seems to be manual, and the extension website offers guidelines for how to label. It urges a conservative attitude, with a set of examples of what is not enough – and what *is* enough – to label someone anti-trans or trans-friendly (Kiran, 2022). The criteria used are based on anti-trans and anti-nonbinary sentiment, and it is consequently targeted at a subset of the challenges related to SOGIESC more broadly.

FIGURE 5.1 Three labeled users on Twitter, the left being labeled trans-friendly while the center and right profiles are marked as trans-hostile.

FIGURE 5.2 Shinigami Eyes user interface.

If an existing label is thought to be wrong, or perhaps in need of updating as referred to in Figure 5.2, the users can simply add new labels, which "will help us [Kiran] improve this extension" (Kiran, 2022). Kiran also states that malicious and fake reports can be immediately overridden locally, while only changes that "pass some trustworthiness criteria (including human validation)" are included in the public dataset used to label accounts for all Shinigami Eyes users.

The browser extension provides an example of how technology enables marginalized and vulnerable groups to take action to protect themselves and create safer spaces online. Having seen the challenges faced by LGBTQ+ people, and trans people in particular, the perceived need to take action online is understandable. In fact, it might be the only way for

these people to feel safe *enough* online to not have to fully withdraw from certain platforms and networks. Seeing how important these social arenas are in many peoples' lives today, this is no small matter (Jenzen, 2017).

However, the extension has also generated controversy related to how such tools potentially change social media as forms of public spheres and could have negative consequences for those being labeled, and we will shortly turn to such objections. First, however, we note that more technical issues are also threatening to derail Shinigami Eyes. For example, Datatilsynet (the Norwegian Data Protection Authority) recently issued a ban on processing after receiving a complaint on the extension from a Norwegian individual marked as trans-hostile by Shinigami Eyes (Datatilsynet, 2022a).

After publishing their intent to impose a ban, Datatilsynet received no replies from the publishers of Shinigami Eyes, several more e-mails from the original complainant, and on June 16, 2022, they proceeded to state that

> Even without a response from Shinigami Eyes we have concluded that we have sufficient information in order to conclude on the legality of the browser extension, i.e. that the processing of personal data in question is in violation of Article 6(1), Article 12(2) and Article 14 GDPR.
>
> (Datatilsynet, 2022a)

The legal details are of less interest in this context than the overarching arguments used to justify the ban. First of all, they argue that marking trans-hostile and trans-friendly people online might constitute protecting people from harm, and consequently that Shinigami Eyes pursues a legitimate interest (Datatilsynet, 2022a). Furthermore, they agree that data processing might be necessary to pursue this legitimate interest. However, they argue that the processing done through the extension constitutes "profiling" according to Article 4(4) GDPR, and that being marked entails potentially significant negative consequences for the ones marked. They can "lose their job, or friendships, and the individual could be the target of hate and mistreatment", and the individuals in question are not notified of being marked or provided any insight into *why* they are labeled (Datatilsynet, 2022a).

This led Datatilsynet to issue a ban, as the potential harms are found weightier than the potential benefits and the strength of the legitimate interest pursued through Shinigami Eyes. They also mention the issue that such extensions can create "chilling effects" online, seemingly mainly noting the chilling effects on those who fear marking, and not on trans people who fear harassment and encountering anti-trans sentiment online if they express themselves and their identities openly (Datatilsynet, 2022a). The potential chilling effects associated with social media filtering as here discussed should not be underplayed, as there are good reasons to be wary of a development where the debate about LBGTQ+ issues, and trans-related issues in general, is fully closed. Our purpose in this chapter, however, is to more thoroughly explore the implications for those being discussed – and how open debate influences them.

Datatilsynet seems to state that individuals have the option not to engage with whomever they want on an individual basis, and indicate that trans people should do so instead

of relying on a "collective decision-making and categorization" which "could strengthen echo chambers found online" (Datatilsynet, 2022a).

In closing, they state that "the legitimate interest pursued by Shinigami Eyes cannot be assessed as one of significant strength or importance", and they consequently impose a ban, the effect of which at the time of writing is unsure. It might be added that the users of the extension do not share traffic with the developers, as the username check is performed locally on each computer. Furthermore, the data stored locally are not accessible in readable format but are represented as a bloom filter (Bloom, 1970), which can be used to test whether a particular account is in the set or not, but not list the entries in the set in a human-readable format (Kiran, 2022).

5.4 DISCUSSION: ONLINE PUBLIC SPHERES AND THE NEED FOR SAFE SPACES

The previous section shows how technology can be used by individuals and groups to change and mediate mainstream media and platforms. The chosen example demonstrates how a particularly vulnerable group can use Shinigami Eyes to make mainstream platforms feel safer and more comfortable, enabling them to potentially continue using and taking part in mainstream social media platforms instead of withdrawing from such platforms completely or having platforms tailored exclusively for such groups, such as Trans Time or similar alternatives (Haimson et al., 2020). While such technologies could indeed be good, and could be called "real trans technology" (Haimson et al., 2021), the benefits of maintaining contact between the trans community and non-trans people seem significant. The positive potential of using technologies similar to Shinigami Eyes is consequently also significant, as it potentially enables marginalized individuals to feel safer and included, and to take part in mainstream arenas and platforms of communication considered important for learning, engaging, and constructing identities (Lucero, 2017). In order to evaluate whether their means of achieving such participation is acceptable, we must begin by recognizing that marginalized individuals can experience online participation as riskier than "more privileged counterparts" (Kruse et al., 2018).

Nevertheless, Shinigami Eyes is subject to significant criticism, and we now proceed to discuss the main potential pitfalls related to the use of such technologies. The main objections are connected to (a) ideals related to public spheres, free and open debate, and (b) the rights of those being labeled by such tools. By and large, the objections mainly relate to SDG 16, and we restrict our analysis to the potentially sustainability-related negative effects of Shinigami Eyes in the following.

5.4.1 Public Spheres and the Free Exchange of Ideas

The main objections to Shinigami Eyes under this heading could be summarized as follows:

1. Labeling individuals as friendly/hostile through a non-universal technology distorts and undermines the idea of social media as Habermasian public spheres.
2. The marketplace for free exchange of ideas is undermined by labeling of individuals instead of arguments and messages.

3. Community-exclusive technologies such as Shinigami Eyes promote the emergence of echo chambers online.

5.4.1.1 Habermasian Public Spheres

The first objection suggests that social media labeling undermines the ideal of an open and universally accessible *public sphere* (Habermas, 1991). A public sphere must be open to all and allows for society to engage in "critical public debate" (Habermas, 1991). The public sphere is a realm of society where individuals and groups can freely partake in a special form of public debate where arguments, and not status or tradition, govern the discourse. The public sphere relates to *political* communication (Bruns & Highfield, 2015; Fuchs, 2015; Kruse et al., 2018) and must be independent of economic and political power with a particular emphasis on the absence of state censorship (Fuchs, 2015). This relates to Hanna Arendt's notion of the public *realm* "as the common world that gathers us together", creates bonds, and makes people interested in common affairs and each other (Arendt, 1998).

An ideal public sphere would be conducive to foster common understanding and the foundations for a functioning deliberative democracy (Kruse et al., 2018). This could consequently contribute to promoting SDG 16 and inclusive political institutions, but by extension it could also arguably contribute to increased awareness of and a reduction of differences (SDG 10) and safer and more inclusive societies (SDG 11), for example. The question, then, is whether Shinigami Eyes undermines the public sphere as an arena for communicative action and deliberative democracy through what could be categorized as "avoidant social media behaviors" such as labeling and potentially blocking/hiding/unfollowing other participants in the sphere in question (Kruse et al., 2018).

First, the public sphere is arguably an ideal that has never existed (Bruns & Highfield, 2015). According to Fuchs (2015), Habermas engages in immanent critique when he describes the public sphere in ways that draws out the shortcomings of capitalist society. Assuming that the ideal of the public sphere currently exists, and that trans people labeling trans-hostile people destroys the public sphere, entails critiquing Shinigami Eyes based on a fantasy. There are multiple *publics*, and these different publics do – and arguably should be able to – follow their own norms and use the affordances provided through extensions and add-ons – and through the platforms themselves – according to their interests and needs (Bruns & Highfield, 2015).

Second, social media has arguably never constituted a public sphere (Kruse et al., 2018). People use social media for a variety of purposes, and many legitimately avoid political topics, seek to avoid harassment, and prefer to use these media as a "happy" place for connecting with like-minded people (Kruse et al., 2018). There is also an important difference between capitalist media and civil society media, and it is the latter that arguably carries the potential for engendering a public sphere free of political *and economic* interests (Fuchs, 2015). Facebook, Twitter, Google, YouTube, and the other media and platforms Shinigami Eyes deals with are all private media, and despite formal equality of access these are arenas where "elites remain elites" (Dagoula, 2019). In Habermas' account elites communicated politically with "ordinary people" as an audience (Bruns & Highfield, 2015),

and while social media is different from earlier mass media in this respect, there are still disparities in power and reach.

Third, power and exclusion matter. The notion of the public sphere is arguably also blind to many forms of marginalization and exclusion (Fraser, 2021; Lunt & Livingstone, 2013). While liberal ideals suggest that universal access and individual's freedoms constitute the foundation for justice and at times equity, both material and non-material inequalities lead to important differences in access to and ownership of the arenas of discourse and organization (Fuchs, 2015; Habermas, 1991; Kruse et al., 2018). The public sphere as an ideal has been a space for "educated, rich men", whereas women, "gays and lesbians" have been relegated to the private sphere (Fuchs, 2015), and "digital gatekeepers" and others with power construct systems in which heteronormative content is prioritized and trans-hostile content is accepted (Jenzen, 2017). If Shinigami Eyes helps those most vulnerable to enter and use mainstream media, rather than relying on their own separate platforms (Haimson et al., 2020), this arguably takes us *towards*, rather than away from, the ideal of the public sphere.

While some are also wary of the chilling effect of Shinigami Eyes (Datatilsynet, 2022a), it seems highly unlikely that this potential chilling effect is of more consequence than the chilling effect that occurs from trans people abandoning mainstream platforms because of unbridled and unmoderated trans-hostility. That one of the most marginalized groups in society is supposedly able to undermine free speech and chill free discourse through an extension that helps them identify individuals that are hostile to them seemingly relies on the liberal notion of the parity and natural equality of individuals. We have shown, however, that this is little more than a fantasy, and that marginalization is both real and must be dealt with in order to achieve sustainable development.

5.4.1.2 Marketplace of Ideas

The second objection is closely related to the first and is based on the idea that the "marketplace of ideas" should be as free as possible to generate a broad set of ideas that allows us to consider all options freely in order to more easily arrive at something close to "truth". The idea is often attributed to John Stuart Mill (1985), who argued that even the "ramblings of madmen" are important inputs in such a process. As seen through the other objections, some fear that Shinigami Eyes can contribute to chilling effects, arguing that the marketplace of ideas becomes less free with such a tool, and that people will refrain from taking part in debates, or will moderate themselves, when fearing that they could be marked as trans-hostile.

We posit that few historical media or arenas can be characterized as simultaneously more accessible and unmoderated than, for example, Twitter. As a medium, then, it could serve as an example of what results from a relatively free marketplace of ideas. Social media platforms have been wary of heavily moderating content on their platforms, fearing both accusations of interfering with free speech, but also to be held liable as editorially responsible for content on the platforms (Bozdag, 2013; Southerton et al., 2021). Rather than making these platforms the incarnation of Habermas' public sphere, this has led to a situation in which a large number of users do not engage in politics on these platforms as they there experience incivility, harassment, lack of privacy, and a general disinterest in seeking communicative action and genuine discourse aimed at becoming better informed about

others and others' perspectives (Kruse et al., 2018). Furthermore, non-moderation does not ensure some form of equity or justice, as non-interference tends to promote the interests of those already privileged (Dagoula, 2019).

5.4.1.3 Arguments, Not People

In a press release accompanying the ban on Shinigami Eyes' processing of data, Datatilsynet's acting director Janne Stang Dahl states that privacy is intended to safeguard a space for free speech, and that "the assessment of the quality of an utterance shall not be dependent on the unknown labels that follow a person across the internet" [our translation] (Datatilsynet, 2022b).

This relates to old ideals of separating an argument from those who make the argument (Sætra & Fosch-Villaronga, 2021). One example is Merton's norms of science, of which *universalism* is of particular relevance. The norm states that within the scientific discourse, characteristics of the individual putting forth an argument or making a knowledge claim should not be considered; theories, discoveries, and ideas should be assessed on what their contents are, and according to accepted principles of the scientific method (Merton, 1942). In his seminal 1942 paper on his norms of science, Merton stakes out the same principles for the exchange of ideas and arguments, in line with the quote from Datatilsynet earlier.

In a discussion of *careful consumption* in media, Edmond (2022) describes how markets are increasingly moralized while morality is marketized. We are currently seeing how individuals are increasingly doing just what Merton warned against, namely evaluating products or messages in light of their creators and the process of creation (Edmond, 2022). Rejecting the debates about wokeness and cancel culture as too simplistic, they describe how individuals in efforts of self-improvement and self-care exercise self-control and increasingly screen and filter content. Trans youth in particular need coping strategies and have found measures for protecting themselves (Jenzen, 2017), such as Shinigami Eyes.

Datatilsynet (2022a) referred to their fear that screening and filtering behavior will lead to echo chambers online – a phenomenon related to filter bubbles – in which people are not exposed to divergent ideas (Sætra, 2021). This connects to the preceding objections, and we here merely refer to how research shows that such phenomena, while theoretically feasible, seem to be exaggerated (Bruns, 2019). Bruns and Highfield (2015) show how the existence of multiple publics does not necessarily lead to echo chambers, and that such publics are partly overlapping and not isolated. Furthermore, in the case of Shinigami Eyes, the question is not only whether it creates echo chambers in a particular medium, such as Twitter, but whether it could actually contribute to a *wider diversity* of users finding the platform bearable, and consequently *countering* echo chamber tendencies.

5.4.2 The Clash of Individual Rights of Protection and Privacy Online

A key reason for Datatilsynet's ban on Shinigami Eyes' processing of data was the purported need to protect those being labeled. They state that being labeled could be stigmatizing and lead to loss of friends, work, etc., that there is no way to respond to being labeled, and that those who are labeled are not informed about why they are labeled or that they are labeled at all (Datatilsynet, 2022a).

These are all good reasons to question Shinigami Eyes, and they must be weighed against the potential benefit of the extension. Much of the criticism is, however, aimed at the technical details of how Shinigami Eyes achieves its objectives. This is not something we explore here, and we also argue that most of these concerns could in theory be solved with technological changes while preserving the main functionality of the extension. For example, if Shinigami Eyes was a registered EU entity which provided reasons and insight into the categorization process and a channel for challenging the categorizations made, many of Datatilsynet's concerns would in fact be moot. We believe that informing individuals of why they are listed and providing a way to appeal categorization would be beneficial, and we assume that the reason this is not done is that Shinigami Eyes is nonprofit and seemingly made and maintained by one or very few individuals.

The first concern earlier, however, would remain. It *is* potentially stigmatizing to be labeled as trans-hostile. Simultaneously, while freedom of speech is an important principle, there are very few who argue that one must be able to utter whatever one likes *and* be free from the social consequences of such actions. Taking Twitter as an example, making a trans-hostile post could easily lead to other users quote retweeting the post and calling the user trans-hostile. Furthermore, it could lead to the person being put on a manually assembled "ban list", for example, put together by individuals concerned with informing other trans people of which accounts to potentially avoid if they do not want to encounter trans-hostile content online. In addition, individuals could of course ban the user directly. Being put on such a list, and being quote retweeted and called trans-hostile, is also stigmatizing, but it is also something that most people would accept as necessary costs associated with other individuals' freedom to refrain from being exposed to other people and what they perceive as harmful content. Freedom of speech does not come with an associated duty for everyone else to listen and pay attention to the speaker. Neither does it come with a duty for others to agree with and not challenge the speaker.

SDG 16 refers to equal access to justice and appeals to the rule of law and transparency are clearly valid when discussing the potential negative aspects of Shinigami Eyes. As we have stated, however, it would be theoretically possible to face all legal requirements and still provide categorization of trans-hostile and friendly people in a manner similar to Shinigami Eyes. The chilling effects and effects on public debate have been discussed earlier, and the remaining concern – that it is stigmatizing to be characterized based on one's utterances – seems to us to be outweighed by the needs of others to be informed about and potentially take precautions against people and messages that are hateful and make them feel unsafe.

5.5 CONCLUSION

When weighing diverging interests as we have done in this chapter, it is imperative that we consider the disparities in power and vulnerability of those involved, and we argue that sustainable development often requires us to side with those least well-off. Agenda 2030 is ripe with such emphasis (Sætra, 2022), and if we must choose between the rights of the privileged and those most vulnerable, the SDGs suggest we *leave no one behind*. The internet is a "crucial lifeline" for many LGBTQ+ youth (Jenzen, 2017), and this should urge us to take their concerns and initiatives to increase resilience seriously.

Some of the key dilemmas we seemingly face when evaluating a tool like Shinigami Eyes are:

1. The need to feel safe versus the desire for unmoderated and unfiltered debates.
2. The desire of a community to share information regarding friends and foes versus an ideal in which there are no groups – only individuals.
3. The need to filter messages by sender due to the massive amount of information available versus the ideal of only evaluating arguments as detached from sender.
4. Untraditional forms of self-protection and resistance from the marginalized versus calls for adherence to rules and established institutions by the privileged.

The public sphere and communicative action are indeed important, but we argue that it seems misdirected to require of the trans community that they spearhead such a development and simultaneously support media in which they are marginalized and encounter massive hostility without taking measures of self-care and self-protection (Edmond, 2022). In any society, some will have more power than others, and the extant structures of media and politics will be better aligned with the needs, interests, and capabilities of some groups of people than others. That this leads to "counter-publics" and fosters resistance through various means and tactics is quite natural (Bruns & Highfield, 2015; Jenzen, 2017), and we believe this is not necessarily something to be combatted.

There are *multiple* public spheres (Bruns & Highfield, 2015), and we should do our best to encourage the participation of the most vulnerable and marginalized in the mainstream media and arenas where politics and culture are debated and shaped. Communication technology has provided new media and arenas, but these are privately owned, largely unmoderated, and not exempt from extant social structures and power. The use of technology to create the conditions for overcoming and navigating such landscapes, like Shinigami Eyes or other tools, demonstrates how individuals and groups can to some degree resist and protect themselves. While such technologies should ideally adhere to prevailing laws, we also argue that prevailing laws could perhaps go further in protecting and promoting the interests and needs of those least well-off, rather than relying on ideals of the free market of ideas and a public sphere – one that has arguably never existed – in ways that arguably protect the interests of those already privileged. As forcefully argued by Ongsupankul (2019), without taking the rights of sexual minorities seriously the prospects for achieving sustainable development on all three dimensions are undermined.

NOTES

1 The definitions related to SOGIESC used in this chapter follow the Yogyakarta principles and the Yogyakarta Principles plus 10 (YP plus 10), found at https://yogyakartaprinciples.org/principles-en/

2 https://twitter.com/CaseyHo/status/1452007901585809411?s=20&t=OgwPxEeQdf3uYbXzV1i3Ag