Academic Year
2023 - 2024

Problematic Content and Online Harm in the Lived Experience of Neurodivergent Social Media Users

Hanne Goor

[20184359]

Master's thesis
Master of Communication Studies

Supervisor
Prof. dr. Sander De Ridder

Co-evaluator
Prof. dr. Steven Malliet

University of Antwerp
Faculty of Social Sciences

Abstract

The widespread use of social media today is marked by concerns of safety and online harm. Previous efforts have explored the manifestation and consequences of online harm. Additionally, as social media platforms mainly address harm through a system of content moderation, scholarship has extensively looked into this side of social media. However, the motivation of platforms to moderate their content is not only marked by preventing harm for their users, but also by judicial and commercial motives. Therefore, the user's lived experience might not be the main guide that platforms base their policy and moderation system on. In this study, the lived experience of the social media user is central. Drawing on conversational in-depth interviews with neurodivergent social media users, this study focusses on gaining insight into how this group experiences problematic content and the potential harms this content might cause. Focusing on neurodiversity as a target group, the needs of this often overlooked minority group are looked into.

Abstract (NL)

Sociale media worden vandaag de dag door heel veel mensen gebruikt. Dit gaat gepaard met bezorgdheden over veiligheid en online harm. Eerder onderzoek heeft al uitgebreid de kenmerken en gevolgen van online harm onderzocht. Aangezien sociale media platformen negatieve effecten voornamelijk proberen tegen te gaan via content moderatie, hebben onderzoekers ook dit aspect van sociale media al uitgebreid besproken. Niettemin is de motivatie van platformen om hun content te modereren niet enkel om hun gebruikers te beschermen, maar willen ze vooral wettelijke en commerciële doeleinden volgen. Dit maakt dat de ervaring van de gebruiker niet de belangrijkste leidraad is voor platformen bij het opstellen van beleid en het opmaken van hun content moderatie systeem. In deze studie staat de ervaring van de gebruiker centraal. Aan de hand van diepte-interviews met neurodiverse sociale media gebruikers, wordt er inzicht verkregen in hoe deze groep problematische content ervaart, alsook de mogelijke negatieve effecten die deze content kan teweegbrengen. Door specifiek te focussen op neurodiversiteit als doelgroep, worden de noden van deze vaak vergeten minderheidsgroep aan het licht gebracht.

Keywords: content moderation, social media, content, online governance, neurodiversity

Wordcount: 11832

Content warning: This paper contains potentially triggering topics such as depression, self-harm, eating disorder, abuse, and suicide.

Contents

Introduction

Social media is an open space that enables connection and free speech. At the same time, these platforms are notorious as spaces for negativity, hate, misinformation and troubling content. Therefore, social media use today is marked by concerns of safety and harm. Consequently, research on social media has extensively been looking into the manifestation and consequences of online harm; the occurrence of something online that has negative consequences in "the real world". Social media platforms mainly address concerns about harm and safety through their policy and through content moderation. Content moderation refers to the actions that platforms undertake to deal with problematic content, behavior or users (Caplan, 2018; Gillespie et al., 2020). In recent years, scholars have been extensively looking into the systems of content moderation of the major platforms.

Regardless of the large amount of research on online harm and content moderation, there are still some areas in which current research might fall short. Mostly, the social media user's lived experience of harm and problematic content is not sufficiently taken into account by social media content moderation practices or by content moderation scholarship. Consequently, what counts as problematic content and the organization of content moderation, is based on the platforms' view and not on the users' needs and experiences. Nonetheless, obtaining insight into user experiences is important, because if we do not know what users find to be harmful, then social media platforms cannot be held accountable for not addressing the harms that certain content might cause its users. Therefore, in my study, I aim to capture the user experience of problematic content and online harm. Moreover, I want to give attention to neurodivergent social media users, as this minority group has been mostly overlooked in previous research. Therefore, this study centers the lived experience and needs of neurodivergent social media users. Drawing on conversational in-depth interviews with neurodivergent social media users, this study focuses on how participants experience problematic content on social media and online harm.

Literature review

Online harm

Online harm is a prevalent issue in everyday social media use. However, it remains rather unclear what this harm exactly entails. When can we say harm has been done? And what does this harm mean for the person who has experienced it? On top of that, how has social media specifically played a role in this harm? Offline, harm is more straightforward to identify; when someone is hit by a car and their leg is broken, the harm is obvious. While on the internet such obvious, immediate harm does not always come forward, thus making it difficult to pinpoint what online harm is. Furthermore, risk does not necessarily coincide with harm. Risk is the possibility of encountering something that might (or might not) cause harm, and being at risk does not always mean an actual harm occurred (Livingstone, 2013). Most research has focused on risk rather than harm; asking people whether they encountered something inappropriate online, but not what the harm in that was, if any. Additionally, harm might be personal, which adds to the difficulty of providing a clear framework for online harm. People might be vulnerable to different things, and some people or groups might be more at risk of harm than others (Xiao et al., 2022). Because of the difficulty to identify online harm, it remains a challenge to measure harm in research. Nonetheless, it is vital for users' voices to be heard and, regardless of the challenges and limitations it poses, this kind of research is needed because it can help to expand understanding on how online harm manifests as well as account for how harm is an individual experience.

Building on the works of Livingstone (2013) and Scheuerman et al. (2021), there are multiple types of online harm that can be identified. A first one is physical harm whereby an individual experiences actual bodily injury. This might be sexual abuse, or self-harm instigated by content promoting self-injury. A second type of online harm is psychological harm. Whereas physical harm might still be a bit more direct and plain to identify, this one gets a bit more obscure, as direct effects of something occurring online are more difficult to determine. Psychological harm can range "from an annoyance (at its least severe) to a stressful or traumatic emotional response (at its most severe)" (Scheuerman et al., 2021). It can also include feeling upset and having a low self-esteem among other things (Livingstone, 2013). A third type of harm is social harm, where one's reputation or social relationships are damaged. Non-consensual spreading of sexual images for example, can lead to serious social consequences (Scheuerman et al., 2021). What these types of online harm have in common is that they have consequences in "the real world", even though being instigated online. Thus, online harm is the occurrence of something online, that has offline consequences for the person who experienced or encountered it. Online harm can be severe

(Scheuerman et al., 2021), and therefore deserves proportionate attention in both social media policy and in research.

Social media content moderation

Social media platforms tend to address online harm through their policy and a system of content moderation. Most social media platforms largely depend on user-generated content (UGC), or content that is created and uploaded by users of the platform. For most social media platforms nowadays, this entails an enormous stream of material that is uploaded every single day. To regulate the vast stream of UGC, platforms draw up rules about what content is allowed to be uploaded and what is prohibited. This is defined in the platform's content policy. These rules are formulated in documents frequently referred to as something along the lines of 'Terms of Service' and 'Community Guidelines'. These are two separate documents; the Terms of Service are stated in more legal terms and serve as a kind of contract between the platform and the user, whereas the Community Guidelines are formulated in a more straightforward way with the objective of informing the user of the platform's rules in a more accessible way (West, 2018). Then, these rules are implemented and enforced on the platform by means of content moderation: a range of actions undertaken by the platform to manage unwanted content, behavior, or users.

There are several actions platforms undertake to deal with problematic content. Some of the most prevalent moderation practices include removal, punitive actions, and reduction. Firstly, removal is maybe the most well-known moderation practice, as it is the one that is most visible to the user. Removal means that inappropriate content gets deleted or users who pose unacceptable behavior get suspended (Gillespie, 2022). Many users might even think that moderation coincides with removal. However, removal is by far not the only way in which platforms deal with problematic content. Another more visible way of moderation are punitive actions against users, to discourage users from posting content that is against the platform's policy. YouTube's strike system is a striking example of such a punitive method. When violating policies, a user receives a strike, which brings about consequences such as not being able to upload for one week, start a livestream, etc. Next to removal and penalties, there are more actions that are highly visible, such as fact-checking labels, warning labels or providing counter-speech (Goldman, 2021). However, moderation can also happen invisibly to the user. This is the case with reduction. Although widely used by many platforms, reduction for a long time has not been debated because it is almost entirely invisible on the platform and not as controversial as removal is. In 2022, Tarleton Gillespie brought "reduction" to the attention as an important part of content moderation, as this practice shapes both how content is presented as well as the public debate on online platforms. Reduction consists of

reducing the visibility of, and limiting the circulation of problematic content (Gillespie, 2022). This is mostly applied to content that is found to be problematic, however not problematic to the extent that is has to be completely removed from the platform: "borderline content" as Gillespie (2022) terms it.

To execute this huge moderation task, platforms utilize a combination of automated and human moderation. Automatic systems screen content, "matching new content with known violations" (Gillespie, 2018), this way identifying prohibited content. Besides automatic screening systems, also users identify prohibited content through flagging. Much of this content (automatically detected and flagged) is then reviewed by either automated systems or human moderators, who will decide whether a reported piece of content is actually in violation of the policy or whether it can remain as it is on the platform. Automated content detection seems like the perfect solution for many of the challenges that platform moderation poses; it provides an efficient answer to the enormous amount of social media content, it reduces the need for human moderators, and it "promise[s] to solve the problem of subjectivity" (Gillespie, 2018). Nonetheless, also automated detection tools are subject to bias, as they are still designed and shaped by humans and are embedded in the same platform policies that are constructed by people. As automated moderation does not suffice, human moderators play an important role, even though content moderation might typically invoke images of algorithms and automated tools. Facebook for example had 15,000 moderation employees in 2020 (Jee, 2020).

Content moderation commonly happens after content is published and after inappropriate content is brought to the platform's attention (as opposed to actively seeking it) (Klonick, 2018). This is in analogy with the fact that most platforms rely heavily on the flagging of content by social media users to moderate their content. Users report content they see and think to be at odds with the Community Guidelines (Gillespie, 2018), after which the platform will assess this reported content. Furthermore, platforms largely apply a reasoning of sameness to content moderation. That is to say, the objective of platforms is to create a standardized idea of what problematic content entails, and to apply this standard uniformly and consistently across the platform (Marshall, 2021). This means that what is being moderated is determined for the platform as a whole, and for the users as one large group. One exception to this is geo-blocking, where specific content is hidden for a certain region in the world; thus, a different norm might be set for a specific region. The aim of this sameness approach is to protect everyone equally and to ensure smooth global communication (Angwin & Grassegger, 2017). This seems like a fair way of setting and enforcing policy; the same rules apply to everyone. However, in reality, this approach might fail to protect everyone equally, as equal treatment does not necessarily lead to equal outcomes. For example, deleting content about self-harm might be benefitting people for whom this is triggering, but can

be damaging for people for whom this recognition is helpful. The reasoning of sameness leans on the idea that problematic content and the moderation practices that follow from it "affect all the communities in the same way" (Marshall, 2021), while in reality they might not. Harm might be very personal and what might be causing one person or group harm, might not necessarily be the same for other people or groups. Tarleton Gillespie (2018) likewise acknowledges that distinct communities within a platform's user base "will have competing values", making it unsuitable to apply the same standard for everyone. Therefore, policies and moderation practices that are more diversified, and also address the systemic inequalities in society, might be more desirable than aiming for an equal treatment across the platform (Marshall, 2021).

The power of social media platforms

In recent years, platform companies have increasingly been scrutinized for the way they deal with unwelcome content. Rising concerns about hate speech and disinformation (Caplan, 2018) as well as increasing public complaints about inadequate governance decisions on part of platforms have led the public, along with academics, journalists and policy-makers to pay more attention to content moderation (Gillespie, 2023). In 2016, the US presidential election vigorously laid bare the issue of misinformation and the impact it has on the public debate and democracy. This event raised urgency and fundamentally redirected the conversation surrounding platform governance and content moderation (Gillespie, 2018). Furthermore, platforms have been receiving criticism on their policy coming from various directions. For example, Facebook's real-name policy has been critiqued by drag queens for discriminating them as they typically use pseudonyms for their stage personas (Gillespie, 2018). Another example is the debate about breastfeeding photos being deleted because they would show too much nudity (Gillespie, 2018). And only recently, in November of 2023, did online platform Omegle, long notorious for child abuse, reach the news when its founder Leif K-Brooks decided to take the platform offline, because of what could be called a loss against the battle of moderation. Omegle kept facing challenges of inappropriate and illegal behavior on the website, causing public scrutiny time and time again. This recent example, and the press coverage it received, shows once again how content moderation and platform governance are receiving more attention and therefore are now being treated with more importance and urgency.

This increased attention for platform moderation has further facilitated the debate concerning the power social media platforms hold in our society (Gillespie et al., 2020). Content moderation is not only a matter of dealing with unwanted content, but also a practice that enables platforms to shape what we see and to organize public debate (Caplan, 2018; Gillespie, 2018). Many moderation practices consist of removing or downplaying certain content or users. This means that

the content flow we see on our social media feeds is not neutral but is actively being regulated by the platform company through a set of rules and (hidden) governance decisions. Additionally, social media platforms carry an immense power in constructing the public debate as they "play an increasingly important civic role as platforms for discourse" (West, 2018), acting as moderators in the democratic debate. Again, they are not just neutral intermediaries, but they actively regulate our everyday speech, deciding who is given a stronger voice, who's voice is kept in the background, which opinions are given more prevalence, etc. Not only do platform companies shape what we see and regulate debate, but they also set the value system for what is acceptable and what is not, as they decide what users are allowed to publish and what is prohibited. This is important because governance decisions are not neutral, but rather they are a reflection of the values of the platform company and the people who set the rules (Gerrard, 2020).

Problematic content and the bad actors of social media

Content moderation practices exist to handle problematic content on platforms. Every platform has its own policy and its own Terms of Service, thus deciding separately what content is problematic and should be moderated and what content is acceptable. Nonetheless, commonalities can be found as there are several categories of content that different platforms point out as problematic. In what follows, I will discuss the most prevalent categories of content that are currently being regarded as problematic, and thus also being focused on in content moderation practices (Arora et al., 2023; Díaz & Hecht-Felella, 2021; Gillespie, 2018; Gillett et al., 2022).

Violence

A first prominent category of problematic content is that of violence. This category covers a whole extent of specific content that is disturbing or extremely graphic. This includes for example "gratuitous images of injuries, fight videos circulated by the instigators, cruelty to animals, deliberate acts of political brutality" (Gillespie, 2018). Also content not necessarily showing violence, but glorifying or instigating violence are usually covered in platforms' inventory of prohibited content. X specifies the latter in its "Platform Use Guidelines": "You may not threaten, incite, glorify, or express desire for violence or harm.".

Harassment and hate speech

The fact that social media platforms have created an open space, where more people than ever can participate in public speech, has also inevitably opened up the floor for

harassment, bullying and hate speech. Hate speech is one of the most difficult types of content to moderate, because of its specificity in terms of geographical and cultural nuances, and the difference in meaning it takes on depending on who is saying it (Díaz & Hecht-Felella, 2021). For example, when a person of color uses the n-word, this might be out of reclaiming the word and out of empowerment, whereas a white person commenting the n-word might be problematic. This nuance is also something that is referenced in Facebook's "Community Standards": "[…] speech, including slurs, that might otherwise violate our standards can be used self-referentially or in an empowering way. Our policies are designed to allow room for these types of speech […].". Even though platforms undertake action against hate speech, there is still a lot of leeway for it to occur as platforms also focus on not interfering with users' freedom of expression (West, 2018).

Self-harm and suicide

Another category of content that is causing platforms concern, is imagery of suicide and self-harm. The moderation of this category of content is a difficult consideration for platforms to make. On the one hand, content about suicide or self-harm might be distressing for people who are struggling with these issues, but on the other this content might be important to users who find support and recognition in it (Gillespie, 2018). In many platforms' policies, the element of glorification or encouragement in content relating to suicide and self-harm is a determining factor in whether the content will be allowed or not. In Tumblr's "Community Guidelines" the following can be found: "Don't post content that actively promotes or glorifies self-harm. This includes content that urges or encourages others to: cut or injure themselves; embrace anorexia, bulimia, or other eating disorders; or commit suicide […].".

Sexual content and nudity

Next, sexual content has long been paid extensive attention in content moderation. What exactly counts as sexual, is defined slightly different across different platforms. Some platforms are more allowing, while others are more restraining. In Instagram's "Community Guidelines", it is defined as follows: "[…] we don't allow nudity on Instagram. This includes photos, videos, and some digitally-created content that show sexual intercourse, genitals, and close-ups of fully-nude buttocks. It also includes some photos of female nipples […].". Generally, nudity and sexual activities are prohibited by many popular social media

platforms and, above all, sexualization of minors is highly focused on and strictly prohibited by most platforms.

Misinformation

Misinformation has been a rising concern in the last years. A recent study, based on newsroom posts made by online platforms, even states that misinformation was the most prevalent category of content that caused concern to the platforms they examined (Gillett et al., 2022). Similar to hate speech, this is a complex category to detect and moderate, because a clear-cut definition of misinformation is difficult to establish. Furthermore, it is rather difficult to decide when moderation of untrue or implausible information is an infringement upon someone's freedom of expression, or when it is a necessary interference. Facebook identifies this difficulty, by saying: "Misinformation is different from other types of speech addressed in our Community Standards because there is no way to articulate a comprehensive list of what is prohibited.".

As to deciding what content is deemed problematic, platforms widely apply a rhetoric of "bad actors" (Gillett et al., 2022). This entails that problematic content is often seen as an issue of "social media villains" posting bad content to the platform to harm innocent users. Consequently, it is mostly content that is illegal or thought of as inherently bad that is deemed problematic and thus also focused on in social media policy and content moderation practices. Of course, it is very much valid that these types of content are seen as problematic and require moderation, however limiting the notion of problematic content to a rhetoric of bad actors possibly causes other types of content that might be harmful to be dismissed. Deciding what is problematic based on the question of what is inherently bad, automatically ignores all types of content that are not necessarily bad nor meant to hurt someone, nevertheless can still be harmful to specific communities of social media users. Therefore, a broadening of what is deemed problematic might be necessary to cover all types of content that might cause people harm.

Platform policy as a reflection of the majority

It is useful not only to look at how content moderation is implemented on social media platforms, but also to ask the question of how these policies are set in place and why, as this reflects the values and interests that go behind them. First of all, it is useful to look into why platforms engage in content moderation. Most obviously, moderation exists to safeguard the users and protect them from seeing appalling content. However, it could be argued that this is the least important reason

for platforms to moderate their content. Another reason is that platform companies are pushed by law initiatives to regulate the speech happening on their infrastructure. Correspondingly, content policies are often embedded in a judicial rhetoric and many things that are prohibited on social media platforms correspond to things that are prohibited by law in "the real world". This is no surprise, as it is important for platforms to comply to existing laws because legislative forces of countries determine whether or not a platform company can offer its service to that country's citizens. Lastly, and probably most importantly, engaging in content moderation helps the platform's commercial ends. Moderating problematic content enables the platform to create a good user experience and will increase the time users spend on the platform, which benefits the advertising gains (Klonick, 2018). Also in attracting advertisers, it is beneficial for platforms to prohibit certain kinds of content, as advertisers might not be willing to be associated with a platform that is full of sexual content for example (West, 2018). Thus, it can be said that platforms' governance decisions are mostly embedded in jurisdictional demands and in the commercial ends of the platform as a profitable company, more so than social justice goals (Roberts, 2016).

Secondly, it is meaningful to look into who's interests platforms policies are a reflection of, or in other words who makes the policies and with whom in mind. Tarleton Gillespie points out that content moderation policies for many of the biggest social media platforms are constructed by employees who are "overwhelmingly white, overwhelmingly male, overwhelmingly educated, overwhelmingly liberal or libertarian, and overwhelmingly technological in skill and worldview" (2018). This means that the norms for what counts as problematic content are set by these people, as "rule-setting reflects the worldview of rule-makers" (Gerrard, 2020). The fact that this group is so convincingly homogenous means that these decisions might not take into account the diversity that is present in the user base of social media platforms. To prove that the norms for what counts as problematic are not only set by the policymakers, some argue that the system of user flagging is a way of taking into account the users' values as to what is problematic and what is acceptable. This might in part be true, but also here it is useful to ask the question of who is mainly doing this flagging and whether this group of frequent flaggers is representative for the user base as a whole. Stefanie Duguay points out: "Much of the time it is a dominant population on a platform that does this – people who are motivated to block and report, sometimes politically motivated or morally motivated […]." (Blunt et al., 2021). Furthermore, because content moderation decisions are deeply embedded in the commercial ends of the platform, moderation is attuned to trying to attract as many users as possible. This is because platforms' revenue is largely based on the times users spend on it. This means that what is deemed problematic is suited to users' speech norms (Klonick, 2018), and preferably the norms of the group with the largest appeal. In consequence, the default standards for deciding what is problematic is based on "white, male, abled, English-speaking,

middle-class US citizens" (Olson et al., 2023), as this is the dominant group in society. Thus, it could be concluded that content policies are a reflection of the people who set these rules, and of the commercial ends of the company. Furthermore, the rules are set with mostly dominant groups in mind, while minority groups are subordinated.

Neurodiversity

Neurodivergent people form a minority group that is often overlooked. Neurodiversity refers to the fact that "people differ in neurocognitive functioning just as they do in many other physiological and psychological aspects" (Ungvarsky, 2023). Simply put there are various ways in which the human brain can function, therefore not everyone's brain functions in the same way. People whose brain functions differently than the established cognitive norm are considered neurodivergent, as opposed to neurotypical people whose cognitive functioning does fall within the dominant norm (Legault et al., 2021). Neurodivergence includes several diagnostic categories such as dyslexia, autism, schizophrenia, ADHD, Tourette syndrome, bipolar disorder, and so on (Ungvarsky, 2023). These diagnostic categories are typically seen as disabilities that are pathological in nature, and are also traditionally researched within that framework (Dwyer, 2022). The neurodiversity approaches offer an alternative approach to researching neurodivergent categories, which does not focus on a medicalized, pathological view of neurodiversity. Instead, this approach suggests that the term neurodiversity simply is a reflection of how the human brain naturally functions in diverse ways (Dwyer, 2022). Accordingly, the neurodiversity movement argues that there is no 'normal' way for the brain to function, and that therefore neurodivergent people are not inferior to neurotypical people as "there is nothing inherently "wrong" about any of these differences" (Ungvarsky, 2023). In my research, I will work within the framework of the neurodiversity approaches, aiming not to pathologize neurodivergent participants or their lived experience, meaning; not solely focus on typical challenges their neurodiversity might cause. Nonetheless, even though there is nothing inherently "wrong" with a neurodivergent functioning of the brain, it is still important to recognize the challenges neurodivergent people might face, not because of their inherent features, but because of "an interaction between [their] own characteristics and their environment" (Dwyer, 2022). The dominant neurocognitive norm might not coincide with their own norms and way of thinking. Thus, being neurodivergent means constantly having to function in a neurotypical society, therefore often putting them in a disadvantaged or minority position.

Neurodiversity is a diversity element that is often overlooked and is often not included amongst other diversity aspects such as gender, skin color, sexuality, etc. Nonetheless, as a neurominority, neurodivergent people hold a disadvantaged position in society, and thus this group

needs to receive consideration in the same way that other minority groups might; neurodiversity should be treated as a diversity element, rather than a disability (Pinchevski & Peters, 2015). Also in social media related research, neurodiversity is often not talked about when looking into the effects that social media might have on various groups. Nevertheless, insights have previously been gained into the social media experience of neurodivergent people, across different disciplines including human-centered computing, human-computer interaction studies, and media studies. This research includes insights into the overall nature of social media use among neurodivergent people (Mazurek, 2013; Wang et al., 2020), as well as media representation of neurodivergent individuals (Johnson & Olson, 2021; Reading, 2018). Furthermore, researchers have examined how social media has an impact on the way neurodivergent people establish social relationships and how they engage in social interactions online (Wang et al., 2020). Also, sensory and design aspects of social media and their relation to neurodiversity have been previously researched (Race et al., 2021; Simpson et al., 2023). These insights are valuable, however most social media-related research into neurodivergence heavily rely on a pathological view on neurodiversity and focus solely on typical characteristics of certain diagnostic categories. It is important to recognize that, by doing so, research might be biased towards a stereotypical view of neurodivergence.

When it comes to content moderation scholarship, many comments have been made about several minority groups, such as for example sex workers, people of color, and LGBTQIA+ members. Here again, neurodiversity seems to be mostly omitted as a diversity element and has not often been talked about in relation to content moderation. A recent study into marginalized groups' ethical concerns about social media has included neurodivergence among race, women, LGBTQIA+, physically disabled people, lower socio-economic status, and the Global South (Olson et al., 2023). This study shows that neurodivergent people, as a group, have high rates of concerns about inappropriate social media content, even sometimes "describing anxiety or depression induced from online content" (Olson et al., 2023). This shows that, when included among other diversity elements and without being pathologized, there are important insights to be gained about the needs of neurodivergent social media users in relation to content moderation and social media policy.

Conclusion and research gap

In conclusion, research surrounding online harm and content moderation has focused on a number of different things. Content moderation scholarship ranges from how platforms decide on what counts as problematic, how their Community Guidelines are enforced on the platform, which methods they use to moderate content, to what motivates platform companies to engage in

content moderation. In my research, I want to further focus on the user's lived experience of problematic content. On the one hand, by connecting online harm research to content moderation scholarship, and on the other by applying a more open, unrestrained view to what counts as problematic content. In this way, I wish to gain wider insight in what content social media users themselves experience to be harmful. In doing so, I aim to explore three areas which current research towards online harm and content moderation might not have sufficiently explored yet.

A first thing that current content moderation scholarship might not sufficiently take into account is the user's lived experience. On the one hand, current content moderation scholarship does not take on an open perspective into what counts as problematic or harmful. Platforms, as well as researchers, widely apply a rhetoric of bad actors to their view of problematic content. Furthermore, platform policy is widely embedded in a judicial sphere. Consequently, problematic content is currently mostly seen as content that is inherently bad, or illegal. However, this view dismisses the potential harm that might be present in other types of content that are not necessarily bad, illegal, nor meant to hurt anyone. Much research already pre-determines what content is problematic and asks about a certain type of such content as opposed to letting the social media user decide for themselves what is problematic to them (Im et al., 2022; Kvardová et al., 2021; Morales, 2023). Therefore, I aim to take on a more open and unrestricted perspective on what content is problematic, in that way following the user's experience rather than the platform's ideas. On the other hand, content moderation scholarship and practices do not take into account online harm insights. Research on online harm has tried to identify what online harm is precisely, and several efforts have been made to capture users' voices in what is harmful to them. Online harm research provides useful insights for content moderation practices, however is not taken into account by content moderation scholarship nor by social media platforms. If content moderation practices would take into account these insights, it could be more focused on the user's experience. Gaining insight into users' experiences is important, because not knowing what is harmful to users results in evidence-free rather than evidence-based platform policy (Livingstone, 2013). This consequently leads to less accountability of social media platforms: if we do not know what is harmful, then platforms do not have the responsibility to address this harm and cannot be held accountable when they do not in fact do so. Therefore, the user experience of problematic content will be central to my thesis.

Secondly, the reasoning of sameness that goes behind content moderation does not properly account for users' individual experiences of harm, and for the fact that harm might be very personal. By using a "one size fits" all method in how content is moderated, the platform decides for the user base as a whole what is problematic, while in reality some things might be problematic for some users while not for others. Lastly, I also want to address the insufficiency with which

neurodiversity has been treated. A good amount of content moderation scholarship has investigated particular groups of social media users, but the group of neurodivergent users has not been sufficiently addressed. By taking neurodivergent individuals' experience of problematic social media content into account, I wish to give this minority group a voice in the construction of platform policy and content moderation practices.

Method

Research questions

The aim of my research was to capture participants' lived experience of problematic content on social media and online harm. I focused on the following research questions:

> Q1: How do neurodivergent social media users experience problematic social media content and online harm?
>
> Q2: What types of content are experienced as problematic?
>
> Q3: What negative consequences do participants experience by seeing problematic content?
>
> Q4: Does being neurodivergent play a role in how problematic content and online harm is experienced?

In-depth interviews

To gain empirical insight into how neurodivergent social media users experience problematic content and online harm, I conducted semi-structured conversational in-depth interviews with this specific target group. This method provides a way to explore individuals' experiences. The interviews took place between April 2 and April 26 2024, both online and face-to-face, and lasted between 20 and 60 minutes. The interviews were conducted in Dutch, the native language of all the participants. While conducting the interviews, I followed a prearranged guideline of questions (see appendix 2), while still leaving room for each participant to share their own experience and ask side questions. The guideline was adapted after the first and second interviews, by each time adding based on new insights I gained during these interviews. To draw up the questions, I partly relied on the frameworks of online harm of Livingstone et al. (2011; 2014) and Scheuerman et al. (2021). I also applied the method of Livingstone et al. of asking an open-ended question about what is experienced to be harmful, before giving any examples, to get participants' unbiased view (2014). The interview questions revolved around three main topics:

1. Problematic content on social media;
2. Online harm;
3. The role of neurodiversity in the experience of social media content and online harm.

Furthermore, I incorporated a practical exercise into the interviews, where participants were presented with different examples of possible social media content and had to assess whether they thought this example would be problematic to them or not. These examples included types of content that are typically seen as problematic (such as violence and sexual content) as well as types of content that are not typically seen as problematic (such as family-related content and dangerous behavior in traffic). I used these examples to spark reflection and conversation about what types of content participants would find harmful and why. Not all participants were presented with exactly the same list of examples, but most of the examples re-occurred between interviewees.

Selection of participants

The interviewees were recruited through a social media post on Facebook, Instagram and TikTok. The post was written in Dutch and informed possible participants of who belonged to the target group and what the study would consist of. People could then sign up to participate through a register form or by sending me a message or e-mail. Participants were selected based on three selection criteria: (1) neurodivergent, (2) social media user, (3) over the age of 18. The criterium of being neurodivergent was solely based on people's own report and experience of being neurodivergent, not on a medical diagnosis. Not focusing on diagnosis in the selection of participants meant respecting people's privacy, as well as being as inclusive as possible, as some people might not have had access to such a medical process and some groups are systemically underdiagnosed. Furthermore, the selection of participants was based on convenience sampling. In total, I conducted 15 interviews. I conducted interviews until saturation was reached. Before the interviews, all participants were informed of the aim of the study and asked for their formal consent for participating in the study and collecting their data, by means of a written form. An overview of the participants is provided in the table below.

Gender	Age	Nationality	Diagnostic category*	Social media use**
Female	21	Belgian	Highly sensitive	Instagram, Facebook, YouTube
Male	20	Belgian	ADHD-C	Instagram, Facebook, LinkedIn, TikTok, Snapchat
Female	29	Belgian	Autism spectrum disorder	Facebook, X, Instagram, Snapchat, WhatsApp

Female	21	Dutch	ADHD	Instagram, Facebook, TikTok, Snapchat, WhatsApp
Female	25	Belgian	Highly sensitive	Instagram, Facebook, TikTok
Female	24	Belgian	Autism spectrum disorder & ADHD	Instagram, Facebook, Reddit, X
Female	22	Belgian	Highly sensitive	Facebook, Instagram, Snapchat, LinkedIn, X, Pinterest, WhatsApp
Female	20	Dutch	Highly sensitive	Instagram, TikTok, WhatsApp, YouTube
Female	28	Belgian	Bipolar disorder & anxiety disorder	Instagram, Facebook, LinkedIn, Pinterest
Female	33	Dutch	ADD & anxiety disorder & highly sensitive	Instagram, TikTok
Female	28	Belgian	ADD	Instagram, Facebook, LinkedIn
Female	25	Belgian	Bipolar disorder & ADHD	Instagram, TikTok, Facebook
Female	23	Belgian	ADHD	Facebook, Instagram, X, TikTok, YouTube
Female	22	Belgian	Highly sensitive	LinkedIn, Facebook, WhatsApp, Snapchat, Instagram
Female	19	Belgian	Autism spectrum disorder	Instagram, Facebook, WhatsApp, TikTok, Snapchat

** I followed people's own description of their diagnostic category, which is why the ADHD terminology is not entirely consistent.*

*** This is a self-report of the platforms that respondents use most often.*

Analysis of the data

In the analysis, I made use of the transcribed interviews. All data was made anonymous before starting the analysis process. I coded the interviews using NVivo 12. The focus was on observing the three main themes of my research: (1) problematic content; (2) online harm; (3) the role of neurodiversity. It is important to recognize that, although all participants are neurodivergent, their answers cannot be directly related to their being neurodivergent. Their experience with problematic content and harm does not have a causal relation to being neurodivergent, but rather captures the way they experience the world and social media by extension. Furthermore, the aim of this study is not to be representative for neurodivergent social media users, but more so to be reflective of the lived experience of this group. Drawing on the obtained data, I aimed to answer the research questions by focusing on recognizing participants' individual experiences while also finding potential commonalities between interviewees.

Results and discussion

What is experienced as problematic content?

As discussed in the literature review, social media platforms and scholars currently apply a rhetoric of "bad actors" (Gillett et al., 2022) as a standard for what is seen to be problematic content. Consequently, it is mostly content that is illegal or thought of as inherently bad that is deemed problematic and thus also focused on in social media policy and content moderation practices. In my research, I tried to focus on what participants experienced to be problematic content, rather than focusing on content that is illegal or inherently bad. By doing so, I found that, indeed, these "inherently bad" types of content are experienced as problematic. However, also other types of content that fall outside of this "bad actors" view, can be experienced as problematic or harmful by the user. Many participants talked about content that is never even discussed in social media policy. Below, I compiled the types of content that participants experienced as problematic, differentiating between content that coincides with the "bad actors" approach and content that does not. These are all examples that participants brought up themselves, without me providing a specific example to reflect on.

"Bad actors" content	Other content
Hate towards a specific group	Political content
Relationship abuse	Someone talking about their pet passing away
Misinformation	Self-help video's for ADHD
(Graphic) images of war	Study accounts
(Graphic) images of wounds or physical accidents	The news
Promoting an eating disorder	'What I eat in a day'-video's
Promoting suicide or self-harm	Hospitals or surgery
Animal cruelty	Workout-content with a shaming undertone
Abusive parents	Toxic positivity (e.g. 'disabled people aren't disabled, they just have a different ability')
	Ghost hunting video's

	Spirituality
	Dolls
	Spiders
	Content with a high energy-level or content that is very busy
	People who portray their life to be perfect
	Shouting
	Negative emotions
	Puppy mills
	People who post their children on social media
	Negative comments
	Testimonies of people who are in a difficult situation

As becomes apparent from this list of types of content, and as many respondents indicated, there was a difference between people in what is experienced as problematic and what not. The respondents specifically noted this when asked the question "Do you think that problematic content entails the same for everyone?". Every single interviewee answered this question with arguing that problematic content can be something different for different people. Some participants also noted that they judged a specific type of content to be problematic or harmful for themselves, but not necessarily for other people. One participant argued: "In essence, I don't think this content [with strong emotions] is problematic, but it can be problematic for me.". Also the other way around, some participants judged content to not have an effect on themselves, but possibly could be harmful for other people. One person noted, for example, that they found content about food interesting to watch, but that for people with an eating disorder the same content might be triggering. Furthermore, many participants argued that their life, their experiences, and their personality all play a crucial role in which content is harmful to them. As this is different for everyone, this also means that different types of content are going to be experienced as problematic. Thus, what problematic content entails, is not the same for every social media user. This ties in with the idea of content moderation not being a 'one size fits all',

as discussed in the literature review. In order to fully take into account the user's lived experience, platforms should thus widen their view of what content is problematic and therefore moderated. Furthermore, moderation practices should be more individualized as problematic content entails different things for different people.

Why is content problematic?

I found many different motivations for participants labeling content as problematic. In what follows, I will discuss several reasons for labeling content as problematic that came forward during the interviews.

Larger concerns

One reason for labeling content as problematic is because of the larger, societal effects that it might have. Many participants recognized the manipulative power of social media and its ability to influence the public debate. Several participants voiced concern about certain content possibly influencing people's opinion. One person, for example, talked about how they are annoyed by some content of political parties or individual politicians, because they try to convince and influence young people. Three interviewees mentioned women-unfriendly ideas being spread on social media. One person specifically argued that this might be problematic for the influence this might have on the ideas of young men. Somebody else talked about people presenting their opinions as facts and trying to influence people in that way, arguing:

> They don't give an objective image of situations. It's just their opinion and they are very firm in that and in that way they influence their audience to think the same, but based on their opinion and not facts. And the fact that it's so strong and reaches such a big audience, makes me a little concerned.

Not only influencing people, but also spreading hatred was seen by some participants as a reason to pinpoint content as problematic. One interviewee talked about how they found content that targets one specific minority group to be problematic, arguing: "Everyone should be able to express their own opinion up to a certain point. But from the moment it becomes hurtful, I think that's a line that's being crossed.". Furthermore, many participants talked about negative, spiteful comments on all kinds of content and how they thought this was problematic. This argument is in line with what social media platforms prohibit according to their policies. However, in the experience of the respondents I interviewed, there still seems to be a lot of leeway to spread hate as multiple participants attested to having come across content or comments that were hateful or offensive.

Furthermore, besides concerns about the effects on people who consume the content, some participants also expressed concern for the people who are seen in the content. When posed with the example of sexual content, one participant said they found it problematic because it could have negative consequences for the person in the image. They were worried that the image might be spread without the person's consent and that "you never know what can be done with it". Somebody else offered the example of influencers who's channels or accounts thrive on the posting of their children. They argued:

> Why do you need to share your whole life together with your children? I think they should still have their privacy too. […] These kids are just so young that they can't make their own choices. Yeah, I don't really think that's normal.

Personal reasons

There are also very personal reasons for why people might find a specific piece of content problematic or harmful. Many participants said they did not like seeing a type of content because it related, in a negative way, to something they had experienced in their own life or to them as a person. Four of the interviewees said that content about happy family relations was problematic to them, because they did not have a good relationship with their family and this content confronted them with that. Somebody else talked about how content of their specific phobias, including hospitals, spiders and dolls, caused them to be anxious or even panic. This participant also recognized that these types of content are not necessarily problematic for other people, but that this is very personal:

> Yeah, I think definitely, things that have to do with my phobias is super personal. I know that other people don't have a problem with these things at all. For example, I also have a very big phobia, it's kind of a funny one, but I have a big phobia of dolls. I can't watch that.

There are several more examples of types of content that came up during the interviews that respondents argued to be problematic for personal reasons. One participant offered the example of content in which people seem to have a perfect life. They said they would rather not see this because they would compare themselves and feel bad about their own life. They argued that it might be entertaining to watch for other people, however for them it had a negative effect, because they will only see what is missing in their own life. This is something that most people might not struggle with, but for this participant, it was a very significant reason to say that this type of content is problematic for them to see.

Another personal motivation to pinpoint content as problematic, is when content aggravates a struggle that someone was already dealing with. Three participants, for example, talked about how some content about food, fitness and eating disorders can trigger struggles they already had surrounding food and body image. Another participant talked about how some content can accelerate getting into a depressive episode. Depressive episodes is something they already struggled with, but because of seeing certain types of content, for example about self-harm, this can trigger or even foster a depressive episode. These examples are all very personal and individualized to the participants' lives and personalities. Therefore, these examples are another attest to the notion of problematic content being individual and different for different people.

Purpose of the content

I found that the purpose of the content is often an indicator of whether something is problematic or not. Many of the interviewees made a distinction between what was problematic, based on the intention behind it or the context in which it was posted. Mainly with content containing violence, risky behavior in traffic, and food this came forward. Several of the participants said they would find content containing violence problematic, however in certain contexts they would find it acceptable or even necessary. For example, in the context of showing a war that is happening, the participants found it important that these images are shown so that we are aware of the cruelties that are going on in other parts of the world. Also with showing risky behavior in traffic, some participants argued that it is necessary sometimes to show this. One interviewee said:

> That is such a double-sided thing. I think it is acceptable to show this to remind people how dangerous it is to drive recklessly. Like they do these days with these promo videos that go 'pay attention when you're behind the wheel' and all that, I think then maybe that's the more positive way of using that content to make it clear. […] I think I could better accept these videos with a positive purpose than those that just want to promote problematic behavior in traffic.

Some interviewees talked about how they did not like seeing content about food, because it could trigger negative thoughts about their own eating habits. However, there was a nuance there, because in some cases they did not find this content problematic. One participant argued:

> I think it depends on the contents. Are you just using it to promote a healthy lifestyle for example, then I wouldn't find it problematic. But if you're doing it to influence people, to kind of almost start developing an eating disorder and if you're promoting a very difficult relationship with food, then I would find it problematic.

These examples make clear that it is not only the contents of what is shown that are important, but also the purpose and the context of what is shown.

Effects and harm of problematic content

As mentioned in the literature review, previous research on online harm has made many efforts to understand the harms that online environments can cause. However, when it comes to content moderation scholarship, harm is never talked about, while preventing harm should be one of the core elements of social media content moderation. To measure harm, I partly relied on the frameworks of Livingstone (2013) and Scheuerman et al. (2021). Livingstone talks about "severity and longevity" as important dimensions of harm (2013) and also Scheuerman et al. see severity as essential in measuring harm (2021). In this section, I analyze the effects problematic content had on the participants and the harm this might have caused.

Which effects can content have?

Firstly, it is important to look at which effects social media content can have. During the interviews, many different effects came forward and some also re-occurred with multiple participants. Below, I sum up all the content-based effects that interviewees talked about and also labeled as a negative or harmful effect. When applicable, I also provide the amount of interviewees that mentioned this effect.

- Feeling sad (4)
- Getting angry (4)
- Feeling stressed (4)
- Adopting negative emotions that are displayed in the content (2)
- Feeling depressed
- Feeling irritated
- Getting anxious (2)
- Doubting yourself
- Panicking (2)
- Getting aggressive
- Getting worked up (3)
- Feeling insecure
- Being overstimulated (2)
- Feeling useless
- Relating things to their own life (2)

- Starting a negative spiral
- Starting a depressive episode
- Experiencing a mood switch

As discussed in the literature review, online harm can be categorized into three types of harm: (1) physical harm; (2) psychological harm; (3) social harm (Livingstone, 2013; Scheuerman et al., 2021). The effects that I found during my research, as becomes apparent in the list above, were mainly psychological harms. Also physical harm can be a part of this, as things like anxiety or depression can also have physical repercussions.

Longevity of the effect

When it comes to the time the effect lasted, I found many different answers. Both between different participants as well as within the same participant. Sometimes, participants indicated that the effects did not linger very long, and it would be forgotten within a couple hours or even right after. In other cases, it was the opposite, and participants said that it did linger for a long time; a couple days to even longer. Whether something has a long-term effect seemed to depend largely on how personal the content was to the person who saw it. The more personal someone could relate to it, the longer it seemed to have a negative effect. One interviewee argued: “It doesn’t linger that long. Also because it’s less my… I don’t have an eating disorder, so I can forget that more easily than when I see something about suicide or self-harm.” Also images or video’s seem to have a longer-standing effect than textual content.

Severity of harm

“Simply recognizing that a behavior *is* harmful is not enough – it is also important to understand *how* harmful the behavior is.” (Scheuerman et al., 2021). Therefore, I not only tried to gain insight into which effects content can have, but also how harmful these effects are believed to be. As mentioned above, there are some pervasive psychological effects that social media content can cause, and they might even be long-lasting. This alone might be considered rather severe. Some participants found it distressing that social media content can have those effects on them. Others, however, argued that they found it an inherent part of social media and therefore found it less severe because it is to be expected. Still, also those participants indicated that, if the option was available, they would actively avoid the types of content they find problematic, in order to avoid the harm it causes them. Only two of the participants said they would not avoid this content and would deal with the consequences. The

fact that people want to avoid problematic content and also the effects it might have, suggests that the effects are severe enough to be taken seriously and to be prevented.

Furthermore, there were some circumstances in which participants indicated the severity was larger. Firstly, the medium of the content seemed to have an effect on the severity of the harm. Several interviewees indicated that seeing something visually made the effects stronger than when they would read a text about the same topic or situation. One person said:

> Yes, those are images that linger too long. Those really have a bigger impact or a more lasting impact than words. Even though I know it's about the same deeds, but when you see it, it is an even harder confrontation.

Secondly, the severity of harm seemed to increase when the content was reality and not fiction. This was not the case for everyone, but some participants indicated this made a big difference to them. For some people, it did not matter whether it was real or not as they indicated they found it problematic in both cases, like this participant argued:

> In both cases, I don't want to see it. That's also why I totally can't stand horror movies, where there's so much blood and wounds and all. In both cases, I would just rather not see it. […] It's mainly about the images.

For other people it did make a big difference as it was easier for them to put things in perspective or shake things off when they knew it was not real. This made the effect less severe for them. One participant said:

> For example with stories about war, that are very traumatic, and in hindsight it turns out to be a true story, then you don't have that 'failsafe' of saying "it's just fiction" to comfort yourself, because it really happened. Which makes it much harder to process those emotions, because you're actually sympathizing and grieving, so to speak, with a person, because they really had to go through that. And with fiction I can put things in perspective much more quickly.

The role of neurodiversity

Content relating to neurodiversity

As already discussed, people can have very personal reasons for why content is experienced to be problematic, or what effect content can have on someone. This is the same for content specifically relating to neurodiversity. Some of the participants talked about how they experience this specific content as people who are neurodivergent. An interviewee, who has ADHD, offered the example of

content that is aimed to provide people with ADHD tips and tricks for how to handle certain difficulties that come with ADHD. They explained these types of videos or posts can be frustrating:

> I have really tried so many methods already. And it can be really annoying, because it really looks like a solution. And sometimes it works. And for me, it often does work, for one or two days. But then I can't get it into my routine. And then it's another lost attempt to try and deal with it. Often, this content is kind of a false promise.

So, besides helping people, this content can also have a negative effect. Another participant talked about how sometimes content about neurodiversity does not reflect their experience of being neurodivergent and that this can also lead to frustration. They found it important that content about neurodiversity actually takes into account the experience of neurodivergent individuals, which they found it does not always. They argued that this leads people to have misconceptions about neurodivergent people, and this had a negative impact on their life. For example, the idea that autistic people use autism as an excuse to not have to work; when social media content reflects this idea, it can be frustrating for autistic people who do not want to be seen in that way. Some participants also mentioned that the comments on content that talks about neurodiversity can be very negative towards neurodivergent people. Several respondents felt that neurodivergent people are target of negative comments and incomprehension online. Belonging to this group, these participants said these comments can affect them negatively. Of course, neurodivergent people will be more sensitive to these types of content, because it is personal to them.

Empathy & sensitivity to others' emotions

Social media content, like any other content, can have an impact on people's emotions. For neurodivergent social media users, this effect might even be stronger and more frequent. Many of the participants indicated that their level of empathy was very high, making some content more impactful. One participant remembered a specific example of content that they empathized with:

> This actually happened only recently, an image of a crying mother in the sand and underneath it said it was a crying mother in Gaza who lost her child. Seeing that, I feel a lot of empathy for that woman and I'm like: "This is real.", and then I easily feel experience feelings of unfairness, like "Why do these people have to go through that?". And there's also a bit of anger, even though that wasn't necessary at all. So yes, that content definitely creates negative feelings.

Especially with highly sensitive participants, it stood out that they felt to be extra sensitive to others' emotions and also indicated this sensitivity showed up with social media content where emotions are

shown or something bad happens to someone. A participant, who is highly sensitive, said for example: "I have seen some testimonies of people who are going through a tough situation, who tell their story. Seeing that, I can really sympathize with them or adopt those feelings.". Most interviewees who talked about empathy and sensitivity to emotions in content, also linked this to the fact that they are neurodivergent. These are of course sensitivities that might show up for social media users who are not neurodivergent, but might possibly be more frequent within the neurodivergent group of social media users. This might make the social media experience of neurodivergent users more intense and more impactful when it comes to emotional effects.

Relating to this, the distinction between fiction and reality seemed to be very important for the emotional effect content can have on neurodivergent individuals. Again, especially with highly sensitive participants this stood out during the interviews. Participants seemed to be able to deal with negative emotions more easily when they knew that the content in question was fictional. When knowing that something was real or something bad really happened to someone, the feeling of empathy was stronger and the emotional effect was way more intense for them.

Aggravating struggles linked to neurodiversity

I found that problematic social media content can trigger some struggles that are particularly linked to neurodivergent diagnostic categories. A participant, who has bipolar disorder, talked about how problematic content can affect their bipolar mood episodes[1]:

> But that's the thing with… I don't know, with me, with my bipolar traits, content can really linger for a really long time. I won't say that I can go from manic to depressive in one second, but that [problematic content] can speed up that switch.

Another participant, who has an anxiety disorder, talked about how problematic social media content can play into their anxiety and trigger anxious feelings about something they already were vulnerable to. Furthermore, they talked about how they have certain phobias that are aggravated by their anxiety disorder and said that seeing content that contains these phobias can be very problematic. They linked their reaction to this content with their anxiety disorder, arguing: "I know that that's something that can affect me way more than someone who doesn't have an anxiety disorder.". As also mentioned above, emotional sensitivity can influence the way problematic content can affect neurodivergent

[1] "Bipolar disorder is a type of mood disorder characterized by episodes of depression alternating with episodes of mania or hypermania. […] hypomanic episodes are characterized by a distinct period of abnormally and persistently elevated mood or increased energy […]." (Ortiz et al., 2018)

individuals, especially highly sensitive individuals. This can cause more intense, and longer-lasting emotional reactions than the same content would for someone who is neurotypical. In other words, problematic content might have a stronger, aggravated effect on neurodivergent social media users and can possibly cause more severe harm than for neurotypical social media users. This is because neurodivergent individuals might be more vulnerable to certain harms because they were already present as traits of their diagnostic category.

How can content moderation take into account the user's experience?

Based on the findings of my study, I can draw up some suggestions on how to handle problematic social media content in a way that better takes into account the experience of neurodivergent social media users. Firstly, it is important to recognize that more types of content than the ones currently focused on, might be experienced as problematic. Ideally, these types of content would also be taken into account when social media platforms moderate their content. Furthermore, I also discussed that there are very personal reasons for why people judge content to be problematic, meaning that the experience of problematic content and harm is highly individual. Social media platforms could adopt their moderation practices to this individuality rather than enforcing a 'one size fits all' policy.

Next, the importance of content warnings and providing context also came forward during the interviews. As discussed, the distinction between content being real or fiction was of importance. Knowing that content was fiction, increased both the longevity as well as the severity of the harm for several participants. Social media platforms could take these kinds of vulnerabilities into account by providing more context for some content. One participant even stated that it would be very helpful for them if platforms clarified whether a piece of content is fiction or not. They argued that it would help them regulate the emotional impact that content can have on them. Furthermore, several participants stressed the importance of trigger warnings or content warnings for them. It gives them a higher sense of agency, because then they actually have the option not to see something before having already seen it. Also for these warnings, providing context is important. One participant mentioned that it would be valuable to provide more context within the content warning, explaining what kind of content is behind the warning.

Lastly, the overall sense of agency is important, as many participants argued that they would like to have more individual power to choose what content they do not want to see. When asked the question: "If there was a button on social media that would allow you to disable a certain type of content to show up on your feed, would you use it?", most participants answered that they would use this option. Mainly because they wanted to avoid the negative effect a certain type of content has on

them. Moreover, giving the user more agency also means increasing the individuality with which content is moderated.

Conclusion

Research on online harm has provided valuable insights into how harm manifests online, as well as consequences and severity of this harm. As social media platforms mainly address this harm by means of their policy and content moderation practices, scholars have also extensively looked into the functioning of content moderation. This has created more transparency of what goes behind content moderation. This scholarship shows that platforms rely on a reasoning of sameness when it comes to enforcing platform policy. This means that the same rules and moderation practices are applied in the same way for everyone. Additionally, content moderation is mainly based on a judiciary and "bad actors" view of problematic content, meaning content which is illegal or inherently bad is regarded as problematic. The way in which social media platforms currently moderate their content might not be sufficient to prevent harm and to protect all users equally. The main problem is that content moderation practices are not based on the user's experience, but more on judiciary influences and commercial goals. In consequence, the way problematic content is currently seen and handled might fail to properly protect the user. Content moderation scholarship has not before addressed these insufficiencies and therefore I aimed to do so in my study. To gain insight into how neurodivergent social media users experience social media and online harm, I conducted semi-structured in-depth interviews with 15 neurodivergent social media users. I focused on finding out (1) which content is problematic, (2) why content is problematic, (3) if harm occurs, (4) how harm occurs.

In my research, I found that problematic content and online harm are both very personal concepts. What is seen as problematic content can be different for everyone. Many types of content that participants labeled as problematic, fall outside of the categories of content that platforms currently see as problematic. I found that also very innocent things, like content about studying or spirituality, can be experienced as problematic. Moreover, the reasons for labeling content as problematic are various, ranging from concerns about manipulation to personal vulnerabilities. Furthermore, the effects that such content can have is wide-ranging; from feeling sad, to panicking, to even feeling depressed. Interviewees attested that some effects can sometimes last for a long time, while in other cases it might be forgotten within a couple of hours. The more personal a piece of content was to someone, the longer the effect tended to last. Furthermore, effects can be severe, as they can lead to pervasive psychological impact and participants claimed to find the effects severe enough to want to avoid them.

As I aimed to capture the experience of neurodivergent social media users specifically, I also gained some insights into struggles that participants found are particular to being neurodivergent. Being sensitive to others' emotions and having a high level of empathy makes the impact of some content more impactful and can make the emotional effect of the content bigger. Furthermore,

problematic content can aggravate struggles that are characteristic of people's neurodivergent diagnostic category, like magnifying anxiety or stimulating depressive episodes. This shows that it is important to include neurodiversity among other diversity elements in online harm research, as significant insights are to be gained about this group's needs.

Taking into account the insights gained in this study, I can make some suggestions on how platforms can better take into account their needs when it comes to policy and content moderation. Firstly, it is vital that the view on what problematic content entails is opened up. Secondly, it could be valuable to enable more agency for the user in what they do and do not want to see and to allow for more individuality in terms of which content is moderated. Thirdly, content warnings and providing context are highly important for some people. Therefore, it would be valuable to focus on these devices and to further extend and sophisticate them.

References

Angwin, J., & Grassegger, H. (2017). Facebook's Secret Censorship Rules Protect White Men From Hate Speech But Not Black Children. *ProPublica*. https://www.propublica.org/article/facebook-hate-speech-censorship-internal-documents-algorithms

Arora, A., Nakov, P., Hardalov, M., Sarwar, S. M., Nayak, V., Dinkov, Y., Zlatkova, D., Dent, K., Bhatawdekar, A., Bouchard, G., & Augenstein, I. (2023). Detecting Harmful Content on Online Platforms: What Platforms Need vs. Where Research Efforts Go. *ACM Computing Surveys*, *56*(3). https://doi.org/10.1145/3603399

Blunt, D., Duguay, S., Gillespie, T., Love, S., & Smith, C. (2021). Deplatforming Sex: a roundtable conversation. *Porn Studies*, *8*(4), 420-438. https://doi.org/10.1080/23268743.2021.2005907

Caplan, R. (2018). *Content or Context Moderation? Artisanal, Community-Reliant, and Industrial Approaches*. https://datasociety.net/library/content-or-context-moderation/

Díaz, Á., & Hecht-Felella, L. (2021). *Double Standards in Social Media Content Moderation*. https://www.brennancenter.org/our-work/research-reports/double-standards-social-media-content-moderation

Dwyer, P. (2022). The Neurodiversity Approach(es): What Are They and What Do They Mean for Researchers? *Human Development*, *66*(2), 73-92. https://doi.org/10.1159/000523723

Facebook. *Facebook Community Standards*. https://transparency.fb.com/policies/community-standards/

Gerrard, Y. (2020). Social media content moderation: six opportunities for feminist intervention. *Feminist Media Studies*, *20*(5), 748-751. https://doi.org/10.1080/14680777.2020.1783807

Gillespie, T. (2018). *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*.

Gillespie, T. (2022). Do Not Recommend? Reduction as a Form of Content Moderation. *Social Media + Society*, *8*(3), 1-13. https://doi.org/10.1177/20563051221117552

Gillespie, T. (2023). The Fact of Content Moderation; Or, Let's Not Solve the Platforms' Problems for Them. *Media and Communication*, *11*(2), 406-409. https://doi.org/10.17645/mac.v11i2.6610

Gillespie, T., Aufderheide, P., Carmi, E., Gerrard, Y., Gorwa, R., Matamoros-Fernández, A., Roberts, S. T., Sinnreich, A., & West, S. M. (2020). Expanding the debate about content moderation: scholarly research agendas for the coming policy debates. *Internet Policy Review*, *9*(4). https://doi.org/10.14763/2020.4.1512

Gillett, R., Stardust, Z., & Burgess, J. (2022). Safety for Whom? Investigating How Platforms Frame and Perform Safety and Harm Interventions. *Social Media + Society*, *8*(4), 20563051221144315. https://doi.org/10.1177/20563051221144315

Goldman, E. (2021). Content Moderation Remedies. *Michigan Techonology Law Review*, *28*(1), 1-59. https://doi.org/10.36645/mtlr.28.1.content

Im, J., Schoenebeck, S., Iriarte, M., Grill, G., Wilkinson, D., Batool, A., Alharbi, R., Funwie, A., Gankhuu, T., Gilbert, E., & Naseem, M. (2022). Women's Perspectives on Harm and Justice after Online Harassment. *Proc. ACM Hum.-Comput. Interact.*, *6*(CSCW2), Article 355. https://doi.org/10.1145/3555775

Instagram. *Community Guidelines*. https://help.instagram.com/477434105621119/?helpref=uf_share

Jee, C. (2020). Facebook needs 30,000 of its own content moderators, says a new report. *MIT Technology Review*. https://www.technologyreview.com/2020/06/08/1002894/facebook-needs-30000-of-its-own-content-moderators-says-a-new-report/

Johnson, M., & Olson, C. J. (2021). *Normalizing Mental Illness and Neurodiversity in Entertainment Media*. Routledge.

Klonick, K. (2018). The New Governors: The People, Rules, and Processes Governing Online Speech. *Harvard Law Review*, *131*(6), 1598-1670.

Kvardová, N., Smahel, D., Machackova, H., & Subrahmanyam, K. (2021). Who Is Exposed to Harmful Online Content? The Role of Risk and Protective Factors Among Czech, Finnish, and Spanish Adolescents. *Journal of Youth and Adolescence*, *50*. https://doi.org/10.1007/s10964-021-01422-2

Legault, M., Bourdon, J.-N., & Poirier, P. (2021). From neurodiversity to neurodivergence: the role of epistemic and cognitive marginalization. *Synthese*, *199*(5), 12843-12868. https://doi.org/10.1007/s11229-021-03356-5

Livingstone, S. (2013). Online risk, harm and vulnerability: reflections on the evidence base for child Internet safety policy. *ZER: Journal of Communication Studies*, *18*(35), 13-28.

Livingstone, S., Haddon, L., Goerzig, A., & Ólafsson, K. (2011). Risks and Safety on the Internet: The Perspective of European Children. Full FINDINGS.

Livingstone, S., Kirwil, L., Ponte, C., & Staksrud, E. (2014). In their own words: What bothers children online? *European Journal of Communication*, *29*, 271-288. https://doi.org/10.1177/0267323114521045

Marshall, B. (2021). Algorithmic misogynoir in content moderation practice. https://www.boell.de/en/2021/06/21/algorithmic-misogynoir-content-moderation-practice

Mazurek, M. O. (2013). Social media use among adults with autism spectrum disorders. *Computers in Human Behavior*, *29*(4), 1709-1714. https://doi.org/10.1016/j.chb.2013.02.004

Morales, E. (2023). Ecologies of Violence on Social Media: An Exploration of Practices, Contexts, and Grammars of Online Harm. *Social Media + Society*, *9*(3), 20563051231196882. https://doi.org/10.1177/20563051231196882

Olson, L., Guzmán, E., & Kunneman, F. (2023). Along the Margins: Marginalized Communities' Ethical Concerns about Social Platforms. https://arxiv.org/abs/2304.08882

Ortiz, A., Bradler, K., & Hintze, A. (2018). Episode forecasting in bipolar disorder: Is energy better than mood? *Bipolar Disorders*, *20*(5), 470-476. https://doi.org/https://doi.org/10.1111/bdi.12603

Pinchevski, A., & Peters, J. D. (2015). Autism and new media: Disability between technology and society. *New Media & Society*, *18*(11), 2507-2523. https://doi.org/10.1177/1461444815594441

Race, L., James, A., Hayward, A., El-Amin, K., Patterson, M. G., & Mershon, T. (2021). *Designing Sensory and Social Tools for Neurodivergent Individuals in Social Media Environments* Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility, Virtual Event, USA. https://doi.org/10.1145/3441852.3476546

Reading, A. (2018). Neurodiversity and Communication Ethics: How Images of Autism Trouble Communication Ethics in the Globital Age. *Cultural Studies Review*, *24*(2), 113-129. https://doi.org/10.5130/csr.v24i2.6040

Roberts, S. T. (2016). Commercial Content Moderation: Digital Laborers' Dirty Work. In *The Intersectional Internet: Race, Sex, Class and Culture Online*. Peter Lang.

Scheuerman, M. K., Jiang, J. A., Fiesler, C., & Brubaker, J. R. (2021). A Framework of Severity for Harmful Content Online. *Proc. ACM Hum.-Comput. Interact.*, *5*(CSCW2), Article 368. https://doi.org/10.1145/3479512

Simpson, E., Dalal, S., & Semaan, B. (2023). "Hey, Can You Add Captions?": The Critical Infrastructuring Practices of Neurodiverse People on TikTok. *Proc. ACM Hum.-Comput. Interact.*, *7*(CSCW1), Article 57. https://doi.org/10.1145/3579490

Tumblr. *Community Guidelines*. https://www.tumblr.com/policy/en/community

Ungvarsky, J. (2023). Neurodiversity. In *Salem Press Encyclopedia*: Salem Press.

Wang, T., Garfield, M., Wisniewski, P., & Page, X. (2020). *Benefits and Challenges for Social Media Users on the Autism Spectrum* Conference Companion Publication of the 2020 on Computer Supported Cooperative Work and Social Computing, Virtual Event, USA. https://doi.org/10.1145/3406865.3418322

West, S. M. (2018). Censored, suspended, shadowbanned: User interpretations of content moderation on social media platforms. *New Media & Society*, *20*(11), 4366-4383. https://doi.org/10.1177/1461444818773059

X. *Rules and policies*. https://help.twitter.com/en/rules-and-policies

Xiao, S., Cheshire, C., & Salehi, N. (2022). *Sensemaking, Support, Safety, Retribution, Transformation: A Restorative Justice Approach to Understanding Adolescents' Needs for Addressing Online Harm* Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, https://doi.org/10.1145/3491102.3517614

Appendix 1 – Verklaring op Eer

Ik, ondergetekende, aanvaard de volgende voorwaarden en bepalingen van deze verklaring:

In het kader van het uitvoeren van mijn masterproef aan de Universiteit Antwerpen (UAntwerpen) binnen de faculteit Sociale Wetenschappen, zal ik toegang krijgen tot (technische en andere) informatie van UAntwerpen en/of derde partijen, in geschreven, elektronische, mondelinge, visuele of eender welke andere vorm, met inbegrip van (maar niet beperkt tot) documenten, kennis, data, tekeningen, foto's, filmmateriaal, modellen en materialen. Deze informatie wordt gezamenlijk met informatie voortkomend uit het door mij uitgevoerde onderzoek beschouwd als 'Vertrouwelijke Informatie'.

Ik zal de Vertrouwelijke Informatie uitsluitend aanwenden voor het uitvoeren van het onderzoek in het kader van mijn studies binnen UAntwerpen. Ik zal:

a) de Vertrouwelijke Informatie voor geen enkele andere doelstelling gebruiken;

b) de Vertrouwelijke Informatie niet zonder voorafgaande schriftelijke toestemming van UAntwerpen op directe of indirecte wijze publiek maken of aan derden bekendmaken.

c) De Vertrouwelijke Informatie noch geheel noch gedeeltelijk reproduceren.

Voor de uitvoering van mijn werk verbind ik mij ertoe om alle onderzoeksdata en ideeën niet vrij te geven tenzij met uitdrukkelijke toestemming van mijn promotor(en).

Na de beëindiging van mijn masterproef zal ik alle verkregen Vertrouwelijke Informatie en kopieën daarvan, die nog in mijn bezit zouden zijn, aan UAntwerpen terugbezorgen.

Naam: Hanne Goor

Adres: Campfortstraat 9 2460 Tielen

Geboortedatum en -plaats : 24 januari 2000 - Turnhout

Datum: 15/08/2024

Handtekening: 

Appendix 2 – Interview guideline

Achtergrond	Naam Leeftijd Geslacht Nationaliteit (Diagnostische categorie)
Inleiding	
Introductie	In dit interview wil ik graag meer te weten komen over jouw sociale media gebruik en jouw ervaring met problematische content. Ik ga jou een aantal vragen stellen over dit onderwerp. Ik zal een audio-opname maken van het gesprek. De resultaten worden anoniem verwerkt en zullen enkel gebruikt worden voor mijn thesis. Je mag op elk moment het gesprek stopzetten of weigeren om een vraag te beantwoorden. Als je akkoord gaat met deze voorwaarden, kunnen we starten.

Inleidende vragen	• Neurodivergente ervaring
	  ○ Wil je meer vertellen over hoe het is om neurodivergent te zijn?
	  ○ Wat betekent het voor jou om neurodivergent te zijn?
	  ▪ Wat versta jij onder neurodiversiteit?
	  ○ Hoe is dit voor jou in je dagelijks leven?
	  ▪ Zijn er dingen die je goed vindt aan neurodivergent zijn?
	  ▪ Brengt dit moeilijkheden met zich mee?
	  ▪ Heb je het gevoel dat je dingen soms anders ervaart of beoordeelt dan anderen?
	  ○ Wil je je diagnostische categorie delen?
	  ▪ Hoe lang weet je dit?
	• Sociale media gebruik
	  ○ Welke sociale media platformen gebruik je?
	  ○ Hoe vaak gebruik je sociale media?
	  ○ Is er specifieke content die je graag bekijkt? (Bv. specifieke mensen die je volgt, specifieke filmpjes die je opzoekt, …)
	• Als je denkt aan problematische content op sociale media, wat is dan het eerste wat in je opkomt?
	  ○ Wat zorgt ervoor dat iets problematisch is? Welk effect moet iets teweeg brengen om problematisch genoemd te worden?

Kernvragen	

Problematische content	• Beoordelen van voorbeelden: Ik overloop een aantal voorbeelden van content die op sociale media zouden kunnen circuleren. Voor elk van deze voorbeelden mag je telkens aangeven of je deze content problematisch vindt, of helemaal niet.
	  ○ Geweld (bv. een video waarin een groepje iemand in elkaar slaat)
	  ○ Mensen die tegen elkaar roepen (bv. in een game-video waarin 2 mensen online aan het gamen zijn en roepen tijdens het spel)
	  ○ Voeding (bv. een filmpje waarin iemand laat zijn wat ze eten op een dag)

	○ Gevaarlijk gedrag in het verkeer (bv. video van een dronken bestuurder die bijna mensen aanrijdt) ○ Seksuele content (bv. een foto waarin iemand naakt te zien is) ○ Afwijzing (bv. een video waarin iemand hun verhaal vertelt over hoe ze thuis zijn moeten weggaan omdat hun vader enkel nog met zijn nieuwe vrouw en nieuwe kinderen wilt wonen) ○ Hevige emoties (bv. iemand is zijn kind verloren en is hier heel verdrietig om en is heel hard aan het huilen) ○ Familie (bv. een video waarin iemand vertelt hoe speciaal hun band is met hun mama en hoe dankbaar ze hier voor zijn) ○ Mentale gezondheid (bv. een filmpje waarin iemand praat over hoe het is om een depressie te hebben) • Kan je je voorvallen herinneren waarbij je content tegenkwam op sociale media die jou stoorde, of een slecht gevoel gaf, of die je liever niet had gezien? ○ Wat was de inhoud van deze content? ○ Was dit een foto, filmpje, tekst, een comment, …? ○ Denk je dat deze content bedoeld was om opzettelijk kwaad te doen? • Hoe ben je deze content tegengekomen? ○ Heb je deze content bewust opgezocht? ○ Kwam je deze content toevallig tegen tijdens het scrollen? ○ Was de afzender een vriend/iemand die je volgt?
Online harm	• Denkend aan het voorval waar we het net over hadden, wat vond je dan storend aan deze content? ○ Denk je dat andere mensen deze content ook storend vinden? Wat maakte deze content voor jou problematisch? • Heb je negatieve gevolgen ervaren door het zien van deze content? ○ Wat was deze negatieve impact? Dit kan zijn: ▪ Fysiek (bv. je zag content van iemand die iets gevaarlijks deed en bent dit daardoor ook gaan uitproberen) ▪ Emotioneel (bv. je zag content over een tragische gebeurtenis en dit beïnvloedde je gemoedstoestand) ▪ Sociaal (bv. iemand plaatste een gemene opmerking bij een post van jou, waardoor je ruzie kreeg met iemand) • Is deze content lang blijven hangen? Heb je hier bv. de rest van de dag nog aan gedacht, of was je dit snel vergeten? • Heeft je eigen leven/dingen die je zelf hebt meegemaakt een invloed gehad op de impact die deze content op jou had? • Heeft content een grotere impact op jou wanneer het om realiteit gaat dan wanneer je weet dat het fictie is? • Hoe ernstig vind je dit negatief gevolg? ○ Zou je het erger vinden om een foto te zien dan tekst? ○ Vind je emotionele gevolgen even ernstig als fysieke gevolgen? ○ Vind je deze negatieve impact ernstig genoeg om te zeggen dat sociale media iets moeten doen om dit te voorkomen? • Vind je dat de gevolgen serieuzer moeten genomen worden wanneer dit een negatieve invloed heeft op veel mensen dan wanneer dit enkel een negatieve invloed heeft op een kleine groep/individuen? • Als er een optie zou bestaan om aan te vinken dat je bepaalde content niet meer zou zien op je feed of minder zou zien, zou je die dan aanvinken?

Neurodiversiteit + sociale media content	• Denk je dat het feit dat je neurodivergent bent een rol speelt bij welke soort content jou stoort?
	  ○ Denk je dat mensen die neurotypisch zijn dit ook storend zouden vinden? Of net niet?
	  ○ Of ervaren zij dit hetzelfde als jij denk je?
	• Denk je dat het feit dat je neurodivergent bent een invloed heeft op de impact die sociale media content op jou heeft?

Slotvragen	
	• Vind je dat er veel problematische content circuleert op sociale media? • Vind je dat dit meer voorkomt op het ene platform dan op het andere? • Denk je dat het voor iedereen hetzelfde is wat ze problematisch vinden? • Vind je dat hier meer aandacht aan besteed zou moeten worden door sociale media platformen? ○ Is het de verantwoordelijkheid van sociale media platformen om ervoor te zorgen dat mensen zo min mogelijk negatieve gevolgen ervaren van de content die op hun platform te vinden is?
Afronding	Dan zijn we aangekomen bij het einde van het interview. Zijn er nog dingen die je graag wil toevoegen? Heb je nog vragen voor mij?

Appendix 3 – Informed consent

Beste deelnemer,

In het kader van mijn masterproef aan de Universiteit Antwerpen, voer ik onderzoek uit naar hoe neurodivergente personen problematische content op sociale media ervaren. Hiervoor voer ik gesprekken die bevragen naar dit onderwerp.

Ik zal je zo dadelijk een aantal vragen stellen over je sociale media gebruik, je ervaring als neurodivergent persoon, en je ervaring met problematische content op sociale media. Dit kan ongeveer 30 minuten tot 90 minuten duren. **Je kan tijdens het interview op elk moment beslissen om een vraag niet te beantwoorden of om het gesprek stop te zetten.**

Tijdens dit interview zal een **audio-opname** gemaakt worden om achteraf de analyse zo nauwgezet mogelijk te kunnen uitvoeren. Deze audio-opname is een noodzakelijke voorwaarde om te kunnen deelnemen aan het onderzoek. **Voor de start van de analyses, maken we onze datasets anoniem** door alle herkenbare gegevens los te koppelen van de antwoorden die je hebt gegeven. Vanaf dan ben je op geen enkele wijze meer herkenbaar in alle datasets en resultaten van deze studie. De geanonimiseerde datasets worden digitaal bewaard tot 10 jaar na het beëindigen van de studie.

Bij deelname aan deze studie ben je tijdens het verzamelen van de data herkenbaar voor onze onderzoekers. **Volgens de GDPR-wetgeving zijn we vereist uw naam en voornaam te registreren**.

Je hebt het recht bestanden waarin jij herkenbaar bent in te kijken zolang deze bestanden bewaard worden. Je kan in deze bestanden enkel gegevens over jezelf inkijken. Het wijzigen of schrappen van gegevens kan uitzonderlijk en als dit het doel van het onderzoek niet in gevaar brengt.

De onderzoekers van deze studie hebben het recht anonieme databestanden (waarin deelnemers voor niemand herkenbaar zijn) **te delen met (inter)nationale collega's** in het kader van (verder, later) wetenschappelijk onderzoek. Deze databestanden zijn niet toegankelijk voor andere partijen, inclusief jezelf of bedrijven/organisaties die niet kaderen binnen wetenschappelijke onderzoeksinstellingen.

Bij verdere vragen of opmerkingen kan je contact opnemen met de **contactpersoon van dit project**:
E-mail: hanne.goor@student.uantwerpen.be Telefonisch: +32 471 22 03 24

Met vragen of opmerkingen inzake het verzamelen van gegevens waarmee je herkenbaar bent kan je tevens contact opnemen met de ***Data Protection Officer*** **of de Privacy dienst van Universiteit Antwerpen**: privacy@uantwerpen.be. Als je denkt dat iemand van de bovenvermelde personen jouw persoonsgegevens niet rechtmatig en volgens de wettelijke vereisten verwerkt, dan heb je recht om klacht in te dienen bij de Gegevensbeschermingsautoriteit (contactgegevens beschikbaar via: https://www.gegevensbeschermingsautoriteit.be/). In geval van klachten raden wij evenwel aan om de Privacy dienst van Universiteit Antwerpen te contacteren (privacy@uantwerpen.be). Vaak kunnen eventuele problemen of misverstanden op deze manier snel en eenvoudig opgelost worden.

Met het bijgevoegde **toestemmingsformulier** vragen we jouw expliciete toestemming om deel te nemen aan het onderzoek en om het verzamelde materiaal te gebruiken voor verder onderzoek of voor opleidingsdoeleinden.

Ik heb de informatie in het inlichtingenformulier gelezen en begrepen en (kruis aan wat van toepassing is):
O ik stem geheel vrijwillig in tot deelname,
O ik wens niet deel te nemen aan deze studie (geen naam en handtekening vereist).

AUDIO-OPNAME
Gelieve bij volgende stellingen te schrappen wat niet past:
Ik ga WEL / NIET akkoord met het maken van een audio-opname,
Ik ga WEL / NIET akkoord met het bewaren van de audio-opname zoals toegelicht.

Naam & voornaam deelnemer : ……

Datum: …….. / …… / ……..

Handtekening deelnemer: