International Journal of Communication 18(2024), 27–47 1932–8036/20240005

Visual Hate Speech and Its Discontents: Young Adult Chinese and Malaysians' Perception of Visual Hate Speech

JAMALUDDIN BIN AZIZ♦
The National University of Malaysia, Malaysia

HOLGER BRIEL
Beijing Normal University–Hong Kong Baptist University, People's Republic of China

As visual elements increasingly dominate communication, we propose to include these as constituting elements that extend hate speech to visual hate speech (VHS). The article explores manifestations of hate speech in an Asian context to understand young people's perceptions and attitudes toward VHS. In September 2021, a convenience sampling technique was used to collect data from young Chinese and Malaysians aged between 19 and 23 years. A total of 26 and 28 undergraduates from a Chinese and a Malaysian university, respectively, were selected. These cohorts were used to further study how much, if at all, their perceptions and attitudes were colored by their respective cultural backgrounds. As an exploration, the respondents were given a simple test with visual stimuli to understand their perceptions and attitudes. Their responses were textually analyzed and thematically organized. The finding suggests that while definitions of VHS vary, it is very much an observable phenomenon and that respondents' perceptions and attitudes not only overlap but also differ in significant ways.

Keywords: visual hate speech, free speech, China, Malaysia, social media, hate speech

Hate speech (henceforth HS) has been in existence for millennia, but its current infamy is derived from the fact that in the digital age, it has become ubiquitous and is pervasive on a much grander scale. Kilvington (2021) explains that "it is now easier than ever before to espouse a hateful message and reach audiences across the world in a matter of seconds. This is like a tsunami of hate, cyber-rippling across countries, causing offence, upset and pain" (p. 257). It has indeed infiltrated and is dominating many mediated discourses at an unprecedented rate, resulting in divisiveness and a threat to the virtual as well as the physical public sphere and its traditional functions of allowing rational continued polylogues. As a worst-case scenario, it facilitates the transition from discourse to violence, be it psychological or physical. If one considers that by mid-2020, the social media platform TikTok (Mandarin: 抖音), owned by Beijing-based ByteDance), had been forced to remove 380,000 posts for the year ("TikTok Removes 380000

Jamaluddin Bin Aziz: jaywalk@ukm.edu.my
Holger Briel: Holgerbriel@uic.edu.cn
Date submitted: 2022-04-16

Videos," 2020) and that Facebook had pulled 22,500,000 posts in the first four months of 2020 alone (Wagner, 2020), with no end in sight, the size of the problem becomes apparent.

Historically, HS has been associated mostly with linguistic expressions. There is a growing number of studies on this topic (e.g., summaries of research in Assimakopoulos, Baider, & Sharon, 2017) and how to approach it, as epitomized by the *Journal of Hate Studies*, launched back in 1992. Since then, more defining texts on the subject, such as Waldron's (2012) *The Harm of Hate Speech* and Udupa, Gagliardone, and Hervick's (2021) *Digital Hate Speech*, have also appeared. As communication technology progresses, an ever-increasing amount of communication is now either in multimodal, textual, and visual forms (e.g., in memes) or outright in the dominant visual mode (e.g., in comic strips and videos) and can be accessed via any cabled or mobile device. As the literature review below reveals, this is a development that has only taken root over the last five to eight years or so, and academia is only slowly stepping up to it. Researchers have only now begun to understand that communication is undergoing a rapid shift and that analyzing it requires a necessary updating of tool sets.

First, this article aims to discuss recent theories of HS in a visual context and delineate ways in which visual hate speech (henceforth VHS) definitions need to be amended to update them to the now prevalent digital social media environment. Second, the article reports on and investigates findings from a 2021 research project conducted at two universities, one in Malaysia and one in China, exploring specific manifestations of HS in an Asian context to better understand young people's perceptions and attitudes toward VHS and to answer the question how cultural backgrounds manifest themselves in these phenomena.

Asia as a region was chosen as the place for the study as it has become the fastest growing and advanced region for the use of mobile phone telephony. China was chosen as it arguably is the largest and most advanced media market in the world. Mobile phone usage is at an all-time high, and even when compared with other Asian countries, Chinese Generation Z uses mobile phone technology more than others (cf. Todorov, 2023). Malaysia was appealing too as it harbors a multiethnic population, with conservative Islam being the most regulated religion in the country. Indeed, Chin (2019) argues that "there are a lot of impressionable young Malays [constitutionally also defined as Muslims] in the community who have become indoctrinated by the racist ideology" (para. 21). Therefore, as VHS is often based on ethnic slurs, we hypothesized that students in both countries could be rather familiar with VHS. While the study does have implications for the Asian region, it does not suggest that all findings would be applicable all over Asia.

Research Methodology

After presenting a discussion of the extant literature on HS and VHS and commenting on some of its limitations, this article explores the perceptions and attitudes of young people toward VHS. Employing a qualitative research approach, a convenience sampling technique was used, and two undergraduate cohorts from a Chinese and a Malaysian university, with 26 and 28 members, respectively, were selected. These two groups of respondents were chosen because they had to attend only online classes due to COVID-19 and were therefore easily accessible. In addition, the respondents came from two different Asian cultures, which made it possible to understand heuristically if the perceptions and attitudes were perhaps (also) colored by their cultural backgrounds. The stimuli survey was constructed using images and short videos found on the Internet and

social media platforms. The participants ranged in age from 19 to 23 years, and the survey was conducted in September 2021. As an exploratory study, the respondents were given a simple test with visual stimuli and a set of survey questions for the purpose of understanding their perceptions and attitudes. The brief stimuli survey was distributed using WhatsApp (Malaysia) and iSpace (China). The respondents were also asked to submit their responses via the same applications to ensure confidentiality and a level playing field. The stimuli survey method was used as it was deemed the least intrusive method to garner participants' responses and also scalable for future larger projects. The project was informed by the following research questions:

RQ1: How do students in Malaysia and China react to VHS?

RQ2: How readily do they recognize VHS? Are there any cultural/gender differences in students' reactions to VHS?

Theoretical Considerations

Most literature on HS and VHS frames the two phenomena as a general problem within a Habermasian public discourse. Especially older models, including legal (e.g., Waldron, 2012) and political ones (e.g., Hare & Weinstein, 2009), refer to pre-digital HS and VHS as prevalent in face-to-face or legacy media situations and provide solutions via rigid legislation. This does not mean that research based on older models is not applicable anymore when online HS and VHS are discussed; however, such models would need to be expanded to remain applicable and retain their pertinence. For instance, the rise of HS should be described not only via sociological parameters based on macroeconomic and political interpretations or the threats of globalization and its discontents but also via the inclusion of the cultural and technical aspects foregrounded by the use of social media in current times.

The former approach was, for instance, applied by Erjavec and Poler (2012) regarding Slovenian news websites and by Gagliardone and colleagues (2016) regarding political campaigns in Ethiopia. But even here, further necessary differentiations must be introduced. A case in point is the application of Western models and findings in countries of the Global South. An example is the research by Kang, Rivé-Lasan, Kim, and Hall (2020), which stresses the fact that HS research has been mostly based on Western theories and examples as it was traditionally there that HS occurred in the early days of social media, but that now Asia, with its very high Internet uptake and, equally, high degree of digitization, should move into the foreground of such studies. In doing so, it is insufficient to merely apply Western models to the Asian virtual mediascapes; rather specific, sociocultural, locational, and technological aspects need to be included. The second approach, focusing on the technical aspects of HS is addressed by studies such as those by Malmasi and Zampieri (2017) and Burnap and Williams (2015), which describe the technologies and algorithms behind platforms' attempts to eradicate HS, which is an ongoing struggle and can be solved by technical means up to a point only.

As HS has been thoroughly treated theoretically and practically (e.g., the comprehensive bibliography in Sponholz, 2018), it is its newest manifestation, VHS, that needs to be researched in a more detailed way. Here, older sociological models need to be extended to comprise newer technological aspects, both on the user and developer/provider side. In the following, we will investigate VHS as a specific subset of HS within the framework of social media.

If HS has a long and hurtful history, VHS is of a newer provenience. Recently, much of the semiotic discourse of VHS has moved to social media and online communication. There, gatekeeping functions are much less pronounced, even frowned on, and their absence arguably contributes to a proliferation of VHS on these platforms. VHS may include images, videos, and memes, with most of them containing multi- or trans-modal communication, and now constitutes a significant part of virtual communication. Our definition of VHS is a broad one here and would also cover cyberbullying, which is typically directed not at a group, but at individuals. Its inclusion is justified as the same psychological mechanisms of attack, exclusion, and ridicule are used to cause harm, irrespective of whether the attack is directed toward a group or an individual.

Increasing problems with online HS can also be traced back to the fact that many of those who use it are "power users," that is, they are very active and thereby potentially very influential in social networks and can therefore easily discriminate symbolically. A study by Ribeiro, Pedro, Yuri, Virgílio, and Wagner (2018) found that typically, "hateful users are 'power users' in the sense that they tweet more, in shorter intervals, favorite [sic] more tweets by other people and follow other users more" (para. 17). Gagliardone and colleagues (2016) found that it was not the persons in power who were using HS, but rather fans and fringe groups, which compensate for their lack of power with poignant and hurtful posts. This does not bode well for the virtual public sphere, and it is therefore imperative that such issues are tackled decisively.

As important as addressing VHS is, the issue with it is not as simple as it might seem. In the United States, for example, the First Amendment protects the right to free speech, and this includes HS. Regularly, the U.S. Supreme Court dismisses cases against the Free Speech Amendment, and only in a small number of cases, when "imminent danger" because of HS can be proven, does it side with a ban on such speech. One recent case was *Matel v. Tam* (2017). In its decision, the U.S. Supreme Court confirmed the precedence of free speech over HS once again. In her book *Hate. Why We Should Resist It With Free Speech, Not Censorship*, Nadine Strossen (2018), former president of the American Civil Liberties Union from 1991 to 2008, defends the U.S. legal position on free speech and counters any legal attempts to reign it in with her solution of "Counterspeech."

By defining the right to free speech very broadly, Strossen (2018) makes an important point, one that could arguably also see the support of Jürgen Habermas and his idea of the public sphere and the ideal speech situation. However, one might equally argue that such a situation is hardly ever given and that especially in social media, this virtual public sphere is in need of more stringent regulation than Strossen's (2018) belief in rational self-regulation warrants. Others agree; for example, Hall (2020), when calling for international laws governing online communication. This is also the viewpoint many national courts and the European Union (EU) have taken, stipulating a more prescriptive approach, strongly declaring the limit of free speech to have been reached when it comes to HS, and asking for discussions on civil liberties to be strengthened (cf. European Commission, 2016).

If HS must be assessed within a band stretching from prohibition to free speech, VHS will problematize things further as it adds a visual element to the equation. Ever since Roland Barthes (1977) in his influential *Image—Music—Text* applied the idea of textual analysis to visual and aural media, visual images have come to the fore in cultural studies. This semiotic decoding would intensify further with the multiplication of images due to regulatory changes in and the rapid development of Internet technologies,

allowing for the ever more facile transmission and display of images on a multitude of screen devices. W. J. T. Mitchell's (1987) famous 1980s' declaration regarding the arrival of the pictorial turn provided an apt framework for this development. While traditional academia at the time would still mostly insist on a disciplinary division between textual (literary) studies and image (film) studies, with the entrance of the Internet and accompanying comprehensive software studies (e.g., by Lev Manovich), such divisions have arguably become outdated as tools from other disciplines would become available, for instance, comics studies. Still somewhat of a niche research area in the 1970s, and then only in the United States, France, and Japan, with the global distribution and shaping force of manga, such studies began to yield important results when it came to the relationship between the text and image, for example, with Greimas and Rastier's (1968) influential and dynamic *carré sémiotique* applied to comics in Groensteen's (1999/2007) *Systéme de la bande dessinée*. Especially when it comes to VHS, and its typical intertwining of text and image, comics studies can assist in providing answers to some of the most vexing aspects of VHS.

When transferred to the realm of HS, this means that for some "interpretative communities" (Fish, 1980), some images do not constitute HS or VHS, whereas for others, they do. In *Is There a Text in the Class?* Stanley Fish (1980) proposes that such interpretive communities are formed by readers sharing interpretive strategies or sets of community assumptions to inscribe a text with meaning. As such, they are creating meaning for/in the text. This is easily applicable to visual material as well. In this regard, it is important to remember though, that the Internet has been very instrumental in creating such "interpretative communities," which today we call bubbles. Fish's (1980) intentions were not to create such bubbles but rather to allow for a multitude of readings, which would then be further discussed in the public realm via interchanges of ideas among these communities. Here, today's situation differs as it seems that the material of the bubble becomes ever less permeable and that the Habermasian public realm in a sense is becoming ever more devoid of interchange between the silos and bubbles of social media.

It is also good to remember here that HS and VHS are global phenomena. In the United States, for instance, many acts of violence are either accompanied by HS posts or are a result thereof (cf. Carlson, 2021). As Carlson explains further, HS also has a silencing effect on the expression and participation of minorities on websites (Carlson, 2021). In Asia, HS came to the fore due to several high-profile suicides associated with cyberbullying. In 2007, South Korean singer and actor Lee Hye-ryeon aka U; Nee committed suicide at age 25, followed by actor Choi Jin-sil in 2008 at 39. Boy band Shinee's singer, Kim Jong-hyun, died by suicide in 2017 at age 27; Taiwanese TV celebrity, Peng Hsin-Yi, known as Cindy Young, committed suicide due to cyberbullying in 2015; in China, Renliang Qiao, a famous actor, committed suicide in 2016; Choi Jin-ri, known as Sulli of girl group f(x), did the same in 2019; Japanese Hana Kimura, a 22-year-old wrestler and reality TV star, committed suicide in 2020; and South Korean actress Song Yoo-jung, committed suicide in 2021.

The issue does persist, and it goes straight to the heart of social media. A case in point is the 2020 TikTok controversy regarding racism, with activists claiming that it is not just platform practices that are racist, but their very algorithms (McCluskey, 2020). TikTok readily agreed and promised to alter the algorithm. Blaming technical faults is, of course, a problematic defense as these technologies and tools have been created by people who, *nolens volens*, bring their own biases to the table.

In 2000, the European Commission Against Racism and Intolerance (ECRI; 2020) published the "ECRI General Policy Recommendation No. 6, on Combating the Dissemination of Racist, Xenophobic and Antisemitic Material via the Internet." The worry of the ECRI at the time was the thin line between speech and violence. When it comes to linguistics, John Searle's (1969) idea of speech consisting of speech acts, already strongly insinuates that language itself is action, it is only a short step to seeing some of these speech acts acted out. Indeed, this is how Charles Lawrence III (1990) saw it when he was writing about the psychological injuries victims of HS would incur. From there it is only a short step to physical violence itself. Mullen and Joshua (2004) saw suicide rates among immigrants rise dramatically after they had been exposed to HS. To be fair, many of them have accepted prescribed codes of conduct, as for instance TikTok did in 2020, when it joined the 2016 European Union Code on Countering Hate Speech. The fact remains though that these are voluntary acts, and that there is little, if any, government regulation of what is posted or denied posting on these social media sites.

Definitions of HS and Their Discontents

While its definitions vary, HS is typically defined as "communications of animosity or disparagement against an individual or a group on account of characteristics such as race, color, national origin, sex, disability, religion, or sexual orientation" (Brown-Sica & Beall, 2008, para. 2). Here, it is worth comparing the various definitions of HS given by large platforms. Such a comparison shows that neither these definitions nor the platforms' varying degrees of code of conduct compliance and balance in approaching HS and their attempts in eradicating HS are uniform. This is further proof that universal definitions of the phenomenon are not easy to come by, if indeed they can exist at all.

Twitter defines HS as content that "promotes violence against or directly attack or threaten other people on the basis of race, ethnicity, national origin, sexual orientation, gender, gender identity, religious affiliation, age, disability, or disease" (Twitter, 2017, para. 1), including implicitly, but not explicitly, images.

TikTok (2023) defines HS in the following way:

> We define hate speech or behavior as content that attacks, threatens, incites violence against, or otherwise dehumanizes an individual or a group on the basis of the following protected attributes: Race, Ethnicity, National origin, Religion, Caste, Sexual orientation, Sex, Gender, Gender identity, Serious disease, Disability, Immigration status. (para. 44)

Instagram (2021a), another social media platform that saw exponential growth over the recent years, stated the following: "Between July and September of last year (2020), we took action on 6.5 million pieces of hate speech on Instagram, including in DMs, 95% of which we found before anyone reported it" (para. 3). Its definition of HS includes "attacks or abuse based on race, ethnicity, national origin, sex, gender, gender identity, sexual orientation, religion, disability or disease" (Instagram 2021b, para. 6).

Facebook has a large segment of its Community Rules dedicated to HS, but when it comes to VHS, there is very little to be found. It states, "We believe that people use their voice and connect more

freely when they don't feel attacked on the basis of who they are" (Facebook, 2023, para. 1) and then defines what "objectionable" content includes, mostly along traditional lines like the ones quoted above. It is worth noting that only "voice" is considered in Facebook's definition of HS. Further down the document, though, it does include imagery when defining what kinds of posts are prohibited: "Dehumanizing speech or imagery in the form of comparisons, generalizations, or unqualified behavioral statements (in written or visual form). (Facebook, 2023, para. 5)." After a detailed listing of posts that are prohibited, Facebook (2023) makes the following statement in regard to the threat of harm, their phrase for the Supreme Court's "imminent danger" formulation:

> Facebook looks at a range of signs to determine whether there is a threat of harm in the content. These include but are not limited to: content that could incite imminent violence or intimidation; whether there is a period of heightened tension such as an election or ongoing conflict; and whether there is a recent history of violence against the targeted protected group. In some cases, we may also consider whether the speaker is a public figure or occupies a position of authority. (para. 10)

Here Facebook clarifies its policy that issues relating to "heightened tension" are criteria for banning certain posts. In this last paragraph, there is no mention made of images.

While these are overall good working definitions, they do not cover specific utterances that might exceed the definitions, perhaps because these are derived from very idiosyncratic situations or because social ideas about what constitutes HS by individuals or members of groups have changed. In TikTok's definition, caste is mentioned, something that by its narrow definition would mostly be irrelevant outside of India. However, it does not question the caste system per se, and it is therefore questionable whether the mere mention of caste would or would not be problematic based on cultural contexts. This might be indicative of a larger issue with such social media sites, that is, that they are willing to quickly intervene, if their business model is in question due to protests by their users or, indeed, government directives and laws, but that they are unwilling to use their unquestionable social power and capital to at least address social ills, including systemic negative Othering.

In general, we found one category that is absent from all the definitions of HS or VHS above, that is, class. An examination of texts on inclusion shows that they usually state that nobody will be discriminated against based on religion, race, or gender, as the abovementioned definitions have exemplified. If these three categories are covered by the platforms' declarations, it is curious that the category of class is absent despite the fact that research has clearly shown that classism is one of the most prevalent prejudices used (Gans, 1996; hooks, 2000; Monroy-Hernández, 2013).

One might speculate that this has to do with the implied audience, which is middle class and has an expendable income. One might even claim that the platforms themselves subscribe to an underlying classist prejudice, trying to attract an economically well-off clientele for their advertisers' sake. Especially in teenage online communication, class is a typical element of VHS, as research has shown (Benn, 2021; Kurth, 2012). This was also commented on by our samples. However, it is absent from all of these definitions. It seems that class has not been considered a pertinent category for HS or VHS lately when it

comes to discussions of attacks or alleviating inequalities. But teenagers, the main target group of social media platforms, are very susceptible to class markers as they seek to find their own identities. Legacy media know this and have been quick to exploit this element. Accordingly, many TV teen dramas have been staged within the rich set, creating desires in teenagers that the commodity industry is only too eager to fulfill but with many others left out and feeling inadequate. A case in point is *Beverly Hills 90210* (Spelling, 1990–2000), an immensely popular Hollywood teen drama famous for its product placements. Although the series was aired during pre-social media times and appeared to defend more "grounded" Midwestern values, as exemplified by the characters of the 16-year-old Minneapolitan twins Brandon and Brenda Walsh (Jason Priestley and Shannon Doherty) against SoCal hedonism (everybody else), it emerged that exclusion of poorer children from schoolyard cliques grew due to the series, which drew severe criticism on this account (e.g., McKinley, 1997). One might even argue that the exclusion of class from recent left-wing criticism of the political status quo has paved the way for populists around the globe and that it is highly advisable to again include class as a relevant marker of exclusion/inclusion.

We stated above that it can be difficult for social media websites to regulate VHS since such expressions can appear out of the blue due to language changes and neologisms that might only be recognizable within a certain time frame and/or place. Thus, in late 2021, under the hashtag #Anti2010, VHS began to appear on TikTok attacking children born in 2010 (Coble, 2021). It is unclear, why this exactly happened, but a hypothesis was put forward that saw its origin in the disrespectful song *Pop it Mania* by YouTube celebrity singer Pink Lily (2022), who, as many saw it, over-eulogized people born in 2010. The result was a massive VHS attack on young children born in 2010. Instagram, Twitch, and TikTok were quick to block the hashtag, but the damage had been done. To be fair to these platforms, this was nothing they could have foreseen, and to their credit, they were quick to address the issue and take down VHS associated with this hashtag. Here, the category of intentionality to hurt would help to qualify an utterance as VHS. But even this criterion might be hard to prove by itself and would most likely require contextualization via the threads it appeared in, a user's posting history, or similar.

Addressing Technical Aspects

Critical sociological, legal, and political approaches to the phenomenon of VHS are well suited to understand and interpret its meteoric rise, but only up to a certain point. They need to be paired with discussions of a technological nature as much of the virtual biotopes such VHS appear in are determined by their technological provenance. If VHS were only to become a social phenomenon due to broadband Internet access and the development of powerful graphic cards, the way companies are now dealing with them is once again dependent on technological means. Detecting 22 million VHS posts and deleting them within 24 hours of their appearance, as the EU requires of social media websites, is not something that can be done manually. Companies have thus begun to develop automated systems that would complete this task. Thus, Burnap and Williams (2015) demonstrated that automated classification features "can be robustly utilized in a statistical model used to forecast the likely spread of cyber hate in a sample of Twitter data" (para. 1). Such technical solutions are now routinely used by all large platforms to detect and delete HS. However, issues with VHS do not have equally facile classifiers at the ready and are more difficult to detect, and such programs based on artificial intelligence (AI), while becoming more precise, are still far from perfect. Multi- or trans-modal modes of inquiry into hate

communication are still under development or in testing phases, both for automated and manual operating systems. As is clear from the sheer number of these posts, it is impossible to eliminate them from social media websites manually. Consequently, media companies employ algorithms to do so. These algorithms themselves have recently come under fire as being racist and thus complicit with HS themselves. In a 2019 study on the usage of African-American English (AAE) on Twitter, Davidson, Debasmita, and Ingmar (2019) found "evidence of systematic racial bias in all datasets" (p. 1).

This is a worrying development as these are automated processes that seem to indicate systemic racism inherent in the very tool supposed to combat such racism. Sap, Dallas, Saadia, Yejin, and Noah (2019) came to a similar conclusion when studying another online AAE language set on Twitter and proposed giving annotators dialect and race priming to reduce racial bias in annotation as such trained systems are significantly less likely to label such tweets as offensive. This is certainly a step in the right direction; however, one might ask what happens when people from outside the circle of AAE speakers read these tweets, would they not perhaps find them offensive? The issue therefore is not one that concerns only one cultural set but the overall Internet populace, and it is an important reminder that technology by itself might not be enough to solve this problem.

When interrogating the approach to VHS, one needs to remember that it is more difficult to detect as multiple data modalities need to be analyzed. On their own, visuals might not fulfill the criteria for VHS, nor indeed the texts used; add them both together though, and they might become VHS very quickly. Kiela, Firooz, and Mohan (2021) used the following example:

> Memes pose an interesting multimodal fusion problem: Consider a sentence like "love the way you smell today" or "look how many people love you." Unimodally, these sentences are harmless, but combine them with an equally harmless image of a skunk or a tumbleweed, and suddenly they become mean.

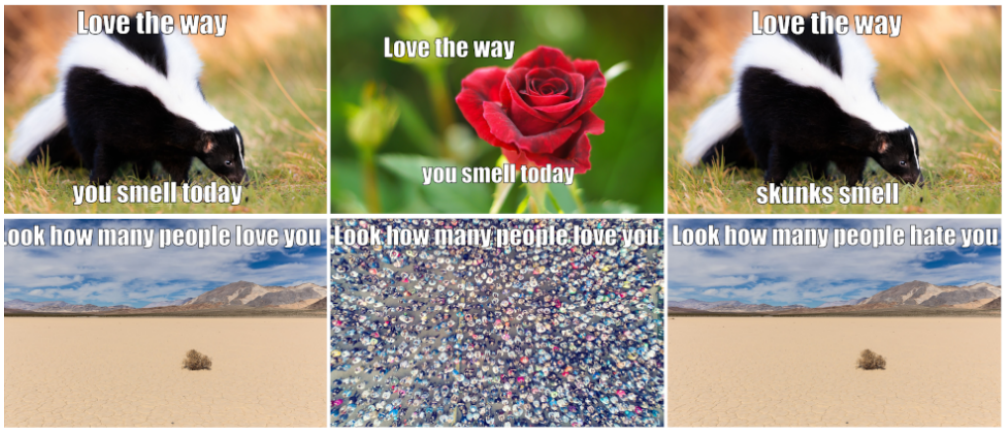

Figure 1. Love the way you smell today (Kiela et al., 2021, pp 1–2).

At best the situation is confusing. It is clear that technological means will have to be applied, otherwise, the sheer quantity of VHS cannot be managed. But the automated process is still far from

perfect, especially when it comes to multimodal approaches. And even if trained AI-based algorithms get better in the future, and issues with biased programming and data sets can be resolved, in one way or another, human supervision needs to be assured. While we do not necessarily subscribe to a U.S.-based system of prioritizing free speech before all else, the question remains who makes the rules deciding what constitutes VHS and what does not. Can this be left up to the social media themselves, whose business model arguably is built on attracting interactions and is less concerned with what these interactions consist of? Or should it be up to governments, who do not necessarily have a record of ideology-free decision making when it comes to public utterances? These are issues that will need to be discussed further in the future, but it is already clear that technology by itself cannot and should not be tasked with giving an (automated) answer.

Discussion of VHS Stimuli Survey

Our approach to the study was a qualitative one. In September 2021, we recruited two sets of volunteer students (N = 54). The first cohort was from China (n = 26), with 21 respondents identifying themselves as female and five as male, all within the 18–22 years age bracket and Year 2 media students at a private South China liberal arts college. The second cohort was from Malaysia (n = 28), with 14 male and 14 female undergraduate students from two public universities, all within the age bracket of 19–26 years. The majority of the respondents were around the ages of 20–21, most of them were Year 2 students; two were Year 3 students. The respondents were asked to fill in a short 18-item survey, which included two memes and two TikTok videos as examples chosen for their applicability to the cohorts:

Figure 2. Meme 1 (Able, 2021).

The meme in Figure 2 shows a Chinese man eating using chopsticks while uttering the words "thats waisis!!" ("That's racist!!") The image makes fun of the Chinese eating culture and the common difficulty among Chinese to pronounce English words correctly (Able, 2021).

Figure 3. Meme 2 (Laframboise, 2018).

The meme in Figure 3 compares two images, suggesting that while plastic straws are banned in California, "immigrant" gangsterism is not (Laframboise, 2018).

Figure 4. TikTok Video 1, Paris Party; screenshot (Craig, 2021).

The video cited in Figure 4 shows a group of French Caucasians at a party organized by a "fashion figure." Each wears a mask of a round Chinese face, with slit eyes, thereby making rude and racist references to a perceived Chinese physiognomy.

Figure 5. TikTok Video 2, Passport Taunt; screenshot (Nas Daily, 2022).

The video in Figure 5 shows famous Israeli-Palestinian TikToker Nuseir Yassin, known as Nas Daily, who was previously not allowed into Malaysia, taunting the viewers with his ability to enter Malaysia using a passport from another country. He does this in a sarcastic tone to show how ridiculous the ban is.

Overall, awareness of VHS was high for both cohorts. We noticed that Malaysian students were more familiar with the theoretical concepts of VHS, perhaps because of the multicultural society in which they were brought up. However, both Chinese and Malaysian students performed similarly when it came to recognizing VHS. The results show that Malaysian students had also encountered VHS slightly more than their Chinese counterparts. Both agreed that this happened overwhelmingly on social media. While these media differed, depending on availability, they were implicated the most in providing a biotope for spreading VHS. Even if the students did not understand the context of a VHS post, all of them reported that they recognized it as VHS and that it made them uncomfortable and feel sad for the target.

From the Chinese cohort, only one respondent reported as being likely to comment on VHS online in a public forum, while three from the Malaysian cohort reported as being likely to do the same. If a VHS had been posted by a friend, the majority in both cohorts reported they would speak to the friend in private, even unfriending him/her if the post was not taken down. An overwhelming majority (52/54) also agreed that VHS was on the rise and that both platforms and governments were not doing enough to combat VHS. Most Chinese respondents opted for a ban on VHS spreaders, while the Malaysians' first choice was to ignore them. Market self-regulation was a distant last choice for both groups. When it came to the memes, the anti-Chinese meme was recognized as VHS by 15 of 26 Chinese and 22 of 28 Malaysians; the latter was a small surprise as we had hypothesized that attacks on one's own nation and/or ethnicity would be more easily identified by in-group members. Malaysia's multiculturalism might have interfered with this assumption and proved it wrong.

The VHS on racism against Latinx in California was recognized as such by 13 of 26 Chinese and 15 of 28 Malaysians. The anti-Chinese TikTok video was recognized by 16 of 26 Chinese and 27 of 28 Malaysians, repeating the strong disdain for anti-Chinese VHS by Malaysian respondents. The anti-Malaysian video was recognized as such by 13 of 26 Chinese and 12 of 28 Malaysians, with the Chinese scoring higher than the targeted Malaysians. Contrary to initial thought, Malaysian students are more inured to such attacks on their ethnicity and country and have grown a proverbial thicker skin than their Chinese counterparts. Conversely, the difference was attributable to the fact that the video did not show VHS as clearly as expected and in future studies, follow-up discussions would help alleviate such differences in understanding intentionality in VHS. Lastly, a majority of the students agreed that further education regarding VHS is necessary and that HS is unfortunately here to stay. One Chinese student summed it up nicely when responding to the question on what individuals can do: "Try to say less uncomfortable words and do less uncomfortable things."

Limitations, Conclusions, and Future Work

Normalization of VHS by the entertainment industry functions less as a solution and more as a part of the problem. The discussion of the meme "Thats waisis!" showed that even "innocent" intentionality does not save moving images from being used in a racist way, thereby being turned against

the original intention of the creator/poster. The whole campaign, in which this meme was used, was well-meaning but was easily co-opted against exactly those people it had been tasked to defend. On top of that, the mainstreaming of such a phrase can be problematic as it takes away from the seriousness of the topic and offers an unhelpful arbitrariness instead, allowing ethnophaulisms to appear exactly where they should not. It speaks in Facebook's favor that it is the only large social media site that expressly includes tongue-in-cheek humor in the midst of possible HS content (see e.g., Swinyard, 2019, for the case of the politicization of Pepe the Frog).

The analysis revealed that most of the participants were aware of VHS and have been affected by it. While they were unwilling to engage with those spreading hatred in online fora, they were very clear that if this happened within their circle of friends, they would intervene. Similar findings to those for the Chinese participants were also recorded for the Malaysian participants. There was a sense of wanting to do the right thing and defend targets of VHS while there were also voices wanting less cancel culture to take place. Results from the survey also show that even little understood non-native VHS leaves a bad taste and is very much disliked.

Equally, the analysis reveals that students were mostly very good at catching HS/VHS, even without understanding the whole context of such posts. While they were quicker in labeling a post pertaining to their culture as VHS/HS they still recognized the structure of VHS/HS even if it did not pertain to their background. A general sense of unease was expressed toward these posts on social media and in semiprivate communities. Future studies should look at these results in detail to arrive at a possibly more detailed semiotic structure of HS/VHS, which would be useful for classification and combating such posts.

Some start-ups have already begun doing so when scouring data for solutions to HS. Thus, Toronto-based hatebase.org aims to help smaller websites and companies that, unlike Facebook or TikTok, do not have vast resources, to detect and alleviate problems with HS, for instance, deleting outright VHS by Incel Groups or the Proud Boys ("No results found. This phrase may be associated with hateful behavior"; TikTok search). However, their approach is still mostly confined to combating text-based HS posts.

Another more pertinent example is the Hateful Memes contest, held by Kiela and colleagues (2021) for Facebook, with 3,937 participants. The results of the contest reveal that most participants were able to identify hate memes and were much better than AI-based VHS detection, which is still in a nascent phase (Doufesh & Briel 2021; regarding the detection of biases in press photographs): "We find that performance relative to humans is still poor, indicating that there is a lot of room for improvement" (Kiela et al., 2021, para. 7). Such AI-based systems are necessarily hampered as they are priority-based and have universalist approaches. Much of the quirkiness of the Internet relies on the ad hoc creativity of its users, which provides additions to or the reworking of existing content. At best, a recent study concluded that AI-based VHS detection systems can therefore at present only play catch-up. Laaksonen, Haapoja, Kinnunen, Nelimarkka, and Pöyhtäri (2020) reported good technical results but also showed that in the process human actors felt ill at ease when it came to technology-only assessments of potential HS. It can be imagined that the more refined the training data sets become, the better such AI-based detection of VHS can become. This is, for instance, the position taken by Cao, Ka-Wei Lee, Ziqing Fan, and Wen-Haw (2021) and their proposed DisMultiHate program. It is unclear

how accurate and focused such systems will become in the future, especially given the fact that the overarching principle of free speech is not supposed to be affected by them, but oftentimes not served as promised.

Lastly, returning to issues of exclusion due to VHS, all the community regulations social media platforms have put in place and are enacting, still suffer from conceptual shortfalls. Some of these are inevitable, such as the arrival of new "exclusion mechanisms" via language or visual displays unforeseeable for even the most adroit automated VHS-detection AI. Others are of a more structural kind. For example, the exclusion of the VHS class category hinted at above. As class is defined differently in different cultures and contexts, we believe that the inclusion of categories addressing rampant classism in VHS detection and their deletion is of paramount importance when updating and adding to the tools designed to fight VHS. When it comes to the visual display of wealth markers, such as designer clothes, mise-en- scènes denoting wealth, exclusion mechanisms are easily put in place and are very effective in leading especially impressionable groups and individuals into doubting their self-worth. We theorized that this exclusion is willfully ignored by social media platforms as their very business model itself is steeped in classism and conspicuous consumption.

As ours was a small-scale study, follow-up studies would be required. For one, studies comparing additional countries in Asia would be helpful to gain a better understanding of how VHS affects Asian young adult communities across the continent. Larger cohorts would also offer more statistically sound results. Finally, follow-up group discussions and focus groups would aid in eliminating misunderstandings regarding individual posts and provide a deeper understanding of how participants reached their conclusions.

Based on the analysis, the fight against VHS cannot be won by single actors alone or against the wishes of social media users. The difficult task is to create tools (legal, cultural, educative, technical) to combat VHS—tools that are stringent enough to address the issues and flexible enough to respond to changing "fads" of VHS. Legislators need to create a legal framework that allows for VHS to be dealt with putatively. Companies need to take the threats of VHS seriously and provide technical and staffing support to push back VHS. Additionally, they would need to include already existing categories of VHS, such as classism, and actively monitor posts for new ones to appear. Lastly, younger Internet users must be educated for them to understand the severity of VHS with potential harm to others and themselves. This is a pan-social endeavor, and only a concerted effort, supported by all stakeholders, can promise success.

References

Able, A. (2021). *"That's racist!"* Know your meme. Retrieved from https://knowyourmeme.com/photos/273566-thats-racist

Assimakopoulos, S., Baider, F. H., & Sharon, M. (2017). *Online hate speech in the European Union—A discourse-analytic perspective*. London, UK: Springer Open.

Barthes, R. (1977). *Image-music-text*. New York, NY: Hill and Wang.

Benn, A. (2021, February 12). *Classism, hate crime and the law commission's consultation paper 250: Lessons from discrimination law*. Retrieved from https://ohrh.law.ox.ac.uk/classism-hate-crime-and-the-law-commissions-consultation-paper-250-lessons-from-discrimination-law/

Brown-Sica, M., & Beall, J. (2008). Library 2.0 and the problem of hate speech. *Electronic Journal of Academic and Special Librarianship, 9*(2). Retrieved from https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1098&context=ejasljournal

Burnap, P., & Williams, M. L. (2015). Cyber hate speech on Twitter: An application of machine classification and statistical modeling for policy and decision making. *Policy and Internet, 7*(2), 223–242. doi:10.1002/poi3.85

Cao, R. R., Ka-Wei Lee, R. C., Ziqing Fan, J. J., & Wen-Haw, C. (2021). *Disentangling hate in online memes*. Association for Computer Machinery. Retrieved from https://arxiv.org/abs/2108.06207

Carlson, C. R. (2021). *Hate speech*. Cambridge, MA: MIT Press. doi:10.7551/mitpress/12491.003.0004

Chin, J. (2019, October 18). *"Setting the stage for race baiting in Malaysia."* The Asian Dialogue Blog. Retrieved from https://theasiadialogue.com/2019/10/18/setting-the-stage-for-race-baiting-in-malaysia/

Coble, S. (2021, September 20). "France condemns #Anti2010 cyber-bullying." *Info-Security Magazine*. Retrieved from https://www.infosecurity-magazine.com/news/france-condemns-anti2010/

Council of Europe-European Commission against Racism and Intolerance. (2020). *Hate speech and violence*. Retrieved from https://www.coe.int/en/web/european-commission-against-racism-and-intolerance/hate-speech-and-violence

Craig, T. C. (2021, April 9). *TikTok Paris party*. TikTok. Retrieved from https://bit.ly/3F8pmEM

Davidson, T., Debasmita, B., & Ingmar, W. (2019, August). Racial bias in hate speech and abusive language detection datasets. In *Proceedings of the Third Workshop on Abusive Language Online* (pp. 25–35). Florence, Italy: Association for Computational Linguistics Anthology. Retrieved from https://bit.ly/2Y1aZ3V

Doufesh, B., & Briel, H. (2021). Ethnocentrism in conflict news coverage: A multimodal framing analysis of the 2018 Gaza protests in The Times of Israel and Al Jazeera. *International Journal of Communication, 15*, 4230–4251.

Erjavec, K., & Poler, M. (2012). "You don't understand, this is a new war!" Analysis of hate speech in news web sites' comments. *Mass Communication & Society, 15*(6), 899–920. doi:10.1080/15205436.2011.619679

European Commission. (2016). *The EU code of conduct on countering illegal hate speech online*. Retrieved from https://ec.europa.eu/newsroom/just/document.cfm?doc_id=42985

Facebook. (2023). *Community standards–hate speech*. Retrieved from https://bit.ly/2Wv3l1q

Fish, S. (1980). *Is there a text in this class? The authority of interpretive communities*. Cambridge, MA: Harvard University Press.

Gagliardone, I., Pohjonen, M., Beyene, Z., Zerai, A., Aynekulu, G., Bekalu, M., . . . Teferra, Z. (2016, May 1). *Mechachal: Online debates and elections in Ethiopia—From hate speech to engagement in social media*. Retrieved from https://ssrn.com/abstract=2831369

Gans, H. (1996). *The war against the poor*. New York, NY: Basic Books.

Greimas, A. J., & Rastier, F. (1968). The interaction of semiotic constraints. *Yale French Studies, 41*, 86–105. doi:10.2307/2929667

Groensteen, T. (2007). *Systéme de la bande dessinée* [System of comics] (B. Beaty & N. Nguyen, Trans). Paris, France: Presses Universitaires de France. Jackson: University of Mississippi Press. (Original work published 1999)

Hall, P. (2020). Dialogues and diversity in Korea, Japan and France. The contribution of international law to hate speech legislation in national and transnational contexts. In M. Kang, M. O. Rivé-Lasan, W. Kim, & P. Hall (Eds.), *Hate speech in Asia and Europe: Beyond hate and fear* (pp. 95–111). London, UK: Routledge. doi:10.4324/9780429264009-9

Hare, I., & Weinstein, J. (Eds.). (2009). *Extreme speech and democracy*. Oxford, UK: Oxford Academic.

hooks, b. (2000). *Where we stand: Class matters*. New York, NY: Routledge.

Instagram. (2021a). *An update on our work to tackle abuse on Instagram*. Retrieved from https://bit.ly/3F7c2jJ

Instagram. (2021b). *Community guidelines FAQs*. Retrieved from https://bit.ly/3urhU2s

Kang, M., Rivé-Lasan, M. O., Kim, W., & Hall, P. (Eds.). (2020). *Hate speech in Asia and Europe: Beyond hate and fear*. London, UK: Routledge. Retrieved from doi:10.4324/9780429264009

Kiela, D., Firooz, H., & Mohan, A. (2021). The hateful memes challenge: Detecting hate speech in multimodal memes. In *4th Conference on Neural Information Processing Systems* (pp. 1–14). Red Hook, NY: Curran Associates Inc. Retrieved from https://proceedings.neurips.cc/paper/2020/file/1b84c4cee2b8b3d823b30e2d604b1878-Paper.pdf

Kilvington, D. (2021). The virtual stages of hate: Using Goffman's work to conceptualize the motivations for online hate. *Media, Culture & Society, 43*(2), 256–271.

Kurth, L. (2012, July 4). *Is classism a hate crime?* Class Action. Retrieved from https://classism.org/classism-hate-crime/

Laaksonen, S.-M., Haapoja, J., Kinnunen, T., Nelimarkka, M., & Pöyhtäri, R. (2020). The datafication of hate: Expectations and challenges in automated hate speech monitoring. *Frontiers in Big Data,* 3, Article 3. doi:10.3389/fdata.2020.00003

Laframboise, D. (2018, August 1). *Plastic straw ban: Humorous meme* [Blog post]. Retrieved from https://nofrakkingconsensus.com/2018/08/01/plastic-straw-ban-humorous-memes/

Lawrence, C. R., III. (1990). If he hollers let him go: Regulating racist speech on campus. *Duke Law Journal, 39*(2), 431–483. doi:10.2307/1372554

Malmasi, S., & Zampieri, M. (2017). Detecting hate speech in social media. In *Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017* (pp. 467–472). Varna, Bulgaria: INCOMA.

McCluskey, M. (2020, July 22). These TikTok creators say they're still being suppressed for posting Black Lives Matter content. *TIME*. Retrieved from https://bit.ly/3il7m0b

McKinley E. G. (1997). *Beverly Hills, 90210: Television, gender, and identity.* Philadelphia: University of Pennsylvania Press.

Mitchell, W. J. T. (1987). *Iconology: Image, text, ideology*. Chicago, IL: University of Chicago Press. doi:10.7208/chicago/9780226148052.001.0001

Monroy-Hernández, A. (2013, April 29). *Classism, accountability, and social media*. Retrieved from https://blogs.harvard.edu/andresmh/2013/04/classism-accountability-and-social-media/

Mullen, B., & Joshua, M. S. (2004). Immigrant suicide rates as a function of ethnophaulisms: Hate speech predicts death. *Psychosomatic Medicine, 66*(3), 343–348. Retrieved from https://pubmed.ncbi.nlm.nih.gov/15184693/

Nas Daily. (2022, December 10). *I finally went to this country* [Video file]. TikTok. Retrieved from https://www.tiktok.com/@nasdaily/video/7175493136016248066?q=nas%20daily%20malaysia&t=1690020311743

Pink Lily. (2022). *Pop it mania* [Video file]. YouTube. Retrieved from https://www.youtube.com/watch?v=3_z4DDc-2uQ

Ribeiro, M. H., Pedro, H. C., Yuri, A. S., Virgílio, A. F. A., & Wagner, M. J. (2018, March 23). Characterizing and detecting hateful users on Twitter. In *Twelfth International AAAI Conference on Web and Social Media* (Vol. 12). Palo Alto, CA: AAAI Press. Retrieved from Https://arxiv.org/abs/1803.08977

Sap, M., Dallas, C., Saadia, G., Yejin, C., & Noah, A. S. (2019, July). The risk of racial bias in hate speech detection. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 1668–1678). Florence, Italy: Association for Computational Linguistics. doi:10.18653/v1/p19-1163

Searle, J. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge, UK: Cambridge University Press.

Spelling, A. (Producer). (1990–2000). *Beverly Hills 90210* [TV Series]. Los Angeles, CA: Fox Network.

Sponholz, L. (2018). *Hate speech in den Massenmedien. Theoretische Grundlagen und empirische Umsetzung* [Hate speech in the mass media. Theoretical foundations and empirical application]. Wiesbaden, Germany: Springer. doi:10.1007/978-3-658-15077-8_2

Strossen, N. (2018). *Hate. Why we should resist it with free speech, not censorship*. Oxford, UK: Oxford University Press.

Swinyard, H. (2019, June 13). Pepe the Frog creator wins $15,000 settlement against Infowars. *The Guardian*. Retrieved from https://bit.ly/3CRYUNI

TikTok. (2023, April). *TikTok community guidelines*. Retrieved from https://www.tiktok.com/community-guidelines/en/?lang%3Den=

TikTok removes 380000 videos in the US violating hate speech policy. (2020, August 31). *South China Morning Post*. Retrieved from https://bit.ly/3D1nQT6

Todorov, G. (2023, May 11). *Gen-Z marketing statistics*. Retrieved from https://thrivemyway.com/gen-z-marketing-stats/

Twitter. (2017). *Hateful conduct*. Retrieved from https://bit.ly/3m6R6ko

Udupa, S., Gagliardone, I., & Hervick, P. (2021). *Digital hate speech: The global conjuncture of extreme speech*. Bloomington: Indiana University Press.

Wagner, K. (2020, August 11). Facebook pulls 22.5 million hate speech posts in a quarter. *Bloomberg*. Retrieved from https://www.bloomberg.com/news/articles/2020-08-11/facebook-pulls-22-5-million-hate-speech-posts-in-second-quarter

Waldron, J. (2012). *The harm of hate speech*. Cambridge, MA: Harvard University Press.

Appendix 1

Questionnaire for the Survey

Visual Hate Speech Survey
September 2021

Survey Information:

This survey is being used for a research project on Visual Hate Speech (VHS). With this project, the two researchers, . . , are trying to ascertain how VHS is understood and how it can be prevented/ reigned in on social media. Your information will be handled with utmost care and your anonymity is guaranteed. Thank you for agreeing to participate!

Biographical information:

Your age: _____
Your gender: _____
Nationality: _____
Please mark/answer the following questions. Taking the survey should not take you longer than 10 minutes.

1. Are you familiar with the concept of Hate Speech?
   Yes No
2. Are you familiar with the concept of VHS?
   Yes No
3. Have you encountered VHS yourself?
   Yes No
4. Where have you encountered VHS? Multiple answers can be given
   TikTok
   Weibo
   QQ
   Other _____
5. How do you recognize VHS? Multiple answers possible
   I feel uncomfortable when seeing it
   I am/identify with the target of the VHS
   I disagree with its message
   I feel sad for its target(s)
   Other: _____
6. What do you do when you encounter it?
   I ignore it
   I ignore it but feel uncomfortable
   I comment on it
   I will not frequent that feed anymore in the future
   Other: _____

7. A friend of yours posts something that could be viewed as hate speech. How do you react?
   I ignore it
   I speak to them in private and ask them to take down the post
   I comment on the post online
   Other: _____
8. Do you think VHS is on the rise?
   Yes No
9. Do you think TikTok and other social media are doing enough to counter VHS?
   Yes No
10. What should social media do to decrease/eradicate VHS?
    Ignore it
    Ban users who post them
    Flag them
    Trust self-regulation
    Other: _____
11. Should the government/the law intervene?
    Yes No
12. Do you think the following GIF meme is VHS? Why (not)?
13. Do you think the following GIF meme is VHS? Why (not)?
14. Please watch the video in the following link: https://www.tiktok.com/@tinachencraig/video/6948993926292425989?is_from_webapp=v1&web_id7009633953573930498=&is_copy_url=1
    Do you think the video you just watched is VHS? Why (not)?
15. Please watch the video in the following link: https://vt.tiktok.com/ZSJx3H2F2/
    Do you think the video you just watched is VHS? Why (not)?
16. What are the dangers of VHS? Please score from 1 (most dangerous) to 6 (least dangerous)

Disinformation (information disorder such as fake news)
Hurting individuals (physical and mental harm)
Creating dangerous rumors
Discriminatory and vexatious stereotyping (Examples: Racial and gender discrimination)
Visual bullying (With the intention of harming one's reputation)
Switches off critical thinking faculties

17. How do you think VHS can be prevented and/or contained?
18. Any other points you would like to raise regarding VHS?