Is Value Sensitive Design Effective in Social Media for User Mental Health?

Sauhee Han

1. What is Value Sensitive Design (VSD)?

1.1. The Background of VSD

> He will not leap in where angels fear to tread, unless he is prepared to accept the punishment of the fallen angels. Neither will he calmly transfer to the machine made in his own image the responsibility for his choice of good and evil, without continuing to accept a full responsibility for that choice (Wiener, 1988, p.184).

In 1954, as an early interest in the ethical implications of technology, Norbert Wiener emphasized the importance of human's control in technology design for the purpose of building and maintaining a just society (Wiener, 1988). A heated discussion around the ethical impact of technology focused on "privacy, ownership and property, physical welfare, freedom from bias, universal usability, autonomy, informed consent, and trust" (Friedman et al., 2013, p.1).

Then, in the 1970s, recognizing how all the stakeholders, including users and the developers, of one technology are impacted differently by one technology due to varying values and priorities, a design approach, called Participatory Design, that highlights a democratization of the technology development process so that all stakeholders' values could be heard was developed (Bødker & Pekkola, 2010).

In the early to mid-1990s, Batya Friedman et al., who developed Value Sensitive Design approach, critiqued the previous design approaches, including Participatory Design, that their application were limited to certain technical implementations and values and thus, argued a necessity for a design approach that can be applied to a broader setting for broader values to

address the expanding scope of technology use (Friedman et al., 2002; Friedman & Hendry, 2019).

1.2. The Definition of VSD

Value Sensitive Design is a design approach to technology that takes stakeholders' values into consideration through the development process via conceptual, empirical, and technical investigations (Friedman et al., 2013). Conceptual investigation deliberates on what value tensions, a conflict among various values (i.e. usability vs. privacy), among stakeholders emerge and how the technology design supports or harms the values. Empirical investigation further evaluates how the stakeholders experience the technology usage based on their value priorities via quantitative or qualitative research. Lastly, technical investigation analyzes the technical limitations and proactively researches the design to support the values recognized in the conceptual and empirical investigations.

Opposed to technological determinism or social determinism, VSD is grounded upon a belief that there is an interactional impact between values from humans and from technology, and thus, co-evolution of technology and the society (Friedman & Hendry, 2019). Its objective is not ethical perfection but rather ethical progress by taking direct and indirect stakeholders into the design's account (Friedman & Hendry, 2019). Another quality of VSD is its multi-lifespan viewpoint that looks forward to resolving social problems by preserving, re-envisioning, and supporting social progress.

2. Why Research on VSD for Social Media Design?

To first define what social media is, Kapoor et al. define social media as a space of user-driven content that enables users to interact with others for any purposes such as personal, business, and political (Kapoor et al., 2018). The user-generated contents are key to social media.

And as anyone could be a part of the platform, there are various stakeholders and thus, various values and value tensions involved.

With numerous value tensions, even with a careful process of balancing the values, one value can be prioritized over other values. For instance, the value tension between usability and user autonomy may affect loss of user autonomy (Briggs & Thomas, 2015). Thus, a work of identifying stakeholders and their values is vital for a more comprehensive understanding of how a design choice that prioritizes one value to another would affect the stakeholders.

Value sensitive design is an approach that addresses the technology's impact with investigations on a comprehensive range of stakeholders and values (Briggs & Thomas, 2015). Hence, as social media is a space with intricate interactions of numerous stakeholders and values, VSD may play a critical role in recognizing them for a more conscious design.

3. Social Media and User Mental Health

3.1. The Relationship between Social Media and User Mental Health

The discussion around whether or not social media use is detrimental to one's mental health is inconclusive as the correlation between social media and mental health is not straightforward (Keles et al., 2020). There are various factors that affect the social media user's mental health such as the user's motivation of social media use and their prior mental health (Faelens et al., 2021; George, 2019). Moreover, the studies present conflicting results on examining the relationship between the aspects of social media usage and the user's mental health. For instance, in the case of the association between the number of followers on Instagram and the user's self-esteem, one study shows a negative correlation, while another research shows no correlation (George, 2019). Additionally, various research shows the positive impact of social media use on users' mental health as it provides them a space for building social capital, a

positive self-presentation, and peer support from social media's mental health communities (Berry et al., 2017; George, 2019). Therefore, it is difficult to claim that social media is the cause of a user's mental health problems.

At the same time, each specific social media platform creates different experiences for each user and thus, has varying mental health effects. For instance, while Instagram membership does not correlate with lower self-esteem and depression, Facebook membership correlates with lower body satisfaction and eating disorders (Faelens et al., 2021). The features of the platforms such as the limitation of post length, the volume of user attraction, and the self-referential or anonymous nature of user accounts affect the quality of the platform interaction and discourse (Chou et al., 2014). Thus, despite that social media use is not parallel to the user's mental health issues, the varying, specific experience the user has in social media affects and may either support or diminishes their well-being (George, 2019; Padín et al., 2021).

Moreover, to further elaborate on how the user's motivation of social media usage and their prior mental health influences their experience with social media, the frequency of using social media is significantly higher among the ones at higher mental risk than the ones at lower risk due to a feedback loop that reinforces their anxiety that activates more selective attention towards relevant contents or activities (Cohen & Blaszczynski, 2015). And passive social media usage, which involves consuming the content rather than actively facilitating online interactions through commenting and posting, would more likely reduce user's well-being than active social media use (Verduyn et al., 2015).

To sum up, the causal relationship between general social media use and mental health is unclear. However, each user's experience in each social platform varies due to the platform design, user's motivation of social media use, and their prior mental health. Therefore, this

finding shows that social media platforms are accountable for user mental health as with certain platforms' design, users' well-being can either be enhanced or diminished.

3.2. The Cause of User Mental Health Problems at the Platform Level

3.2.1. Economic

Attention economy is an economic model that treats users' attention as a product to fulfill their business clients' advertising demands. The current business model of social media companies is governed by the attention economy as their profit relies heavily on target advertising (Frank, 2021; Gillespie, 2010). Driven by financial incentive, since there is no additional cost caused by taking new user accounts, this business model maximizes profit by not charging the users for a subscription fee but rather garnering more zero-cost users to their platform and harness their engagement as a profitable product (Bhargava & Velasquez, 2021; Frank, 2021). Therefore, to attract more users and effectively profit from their attention, social media companies design their platform to increase user engagement and their screen time.

Raising users' screen time requires the understanding of their patterns of attention. From the leaked Facebook report (Lewis, 2017), it has been revealed that their algorithm is designed to render adequate feedback based on its identification of user's insecurity, loneliness, or boredom to increase user engagement with the platform (Lewis, 2017). LinkedIn social networking feature that exploits users' need for social reciprocity (Lewis, 2017), YouTube and Netflix autoplay feature that lessens the users' autonomous decision-making (Lewis, 2017), Snapchat's streak message that is addictive (Griffiths, 2018; Lewis, 2017), Twitter's pull-to-refresh mechanism, which refreshes the page with a single finger movement of pulling and releasing the screen, that hinders user's autonomous decision-making all target the goal of making the users engage longer with the platform (Bhargava & Velasquez, 2021; Lewis, 2017).

It is critical to note that not all designs' intent was to manipulate users to increase their screen time. In case of pull-to-refresh mechanism, which is now ubiquitously used in online platforms for its success of boosting users' engagement time, its first developer, Loren Brichter, explained that it started as a small, "cute and clever" fix to substitute the "refresh" button in the platform that has limited amount of space on a screen (Lewis, 2017).

Because the economic model of many social media platforms aim for profit, their design's success is significantly defined by its capability to escalate users' engagement (Bhargava & Velasquez, 2021). Even when the initial intent of the designs was not manipulative, the designs are part of that economic model. To encapsulate, the current objective of the social media business model is difficult to put users' well-being as the end goal without considering the company's profit, and it affects the objective of design choices, which is to increase the user screen time.

3.2.2. Legal

There is a lack of regulations that check social media company's content moderation decisions and protect social media users' well-being from the company. The content moderation of social media is primarily governed by social media company's private moderation policies rather than by federal law due to the First Amendment's Free Speech Clause and Section 230 of the Communications Decency Act of 1996 (CDA) (Brannon, 2019). By the First Amendment, a social media company's decision on content moderation is regarded as its free speech and thus, protected when it is considered as a private actor sufficiently distant from the government (Brannon, 2019). And by Section 230 of CDA, which separates social media company from information content provider on its platform and does not hold the social media company liable for its moderation decisions if they are done in "good faith," social media company is provided

with more immunity to potential lawsuits charged against them for the contents presented on its platform (Brannon, 2019).

3.2.3. Political: The Intersection of Economic and Legal Motives

Social media companies strategically frame their service as a technology company that provides users an egalitarian platform to create and distribute content rather than a media company that actively curates and publishes content (Gillespie, 2010). However, their role as a content moderator that designs how to present the content, their intertwined role as a content creator such as YouTube original video production, and their significant impact in shaping public discourse contradict their upheld position as a purely technology company (Napoli, 2019).

In spite of such contradictions, social media companies vehemently advocate their current framing because of economic and legal advantages. The first economic benefit is that the label as a technology company is more appealing to the current investment community than the label as a media company (Napoli, 2019). Secondly, such framing of it as a "platform" attracts both users and advertisers as a space for freedom and opportunity (Gillespie, 2010). Legally, such a frame benefits the company by minimizing their liability for illegal or inappropriate contents that users create (Gillespie, 2010). And by emphasizing their service as a facilitation of democratic discourse, they lobby policies that benefit them (Gillespie, 2010).

Thus, knowing that their self-presentation is highly motivated by economic and legal benefits, their framing as a neutral platform for users is not comprehensive to its role and impact. Hence, the user's expectation of social media to be a neutral platform, where information spurs freely and democratically, is misleading. And social media companies do not manifest their strategic framework to the users because that was their intended expectation, which shows the unbalanced power between the social media company and the users. And the power imbalance

of the company and the users causes a lack of user privacy and autonomy and user's dependence on a provider (Datta et al., 2010).

3.2.4. Technical

Even if the social media company wants to impose a guideline in consideration of users' well-being, the current advancement of technology is incapable of perfect exercise of that moderation. When Facebook was held accountable for unchecked circulation of misinformation about Covid vaccines, it has made various efforts such as removing or hiding more than 185 million false Covid contents, supporting users with a pre-written vaccine fact option they can append to their comment, and adding a feature to turn off the comment section to address the adverse comments (Schechner et al., 2021). While these efforts contributed to intervening the spread of misinformation, its moderation system was not developed enough to detect all the nuances of vaccine hesitant comments and comments with incorrect grammar or non-existent words (Schechner et al., 2021). Therefore, the limitations of technology hinders reinforcing guidelines for users' well-being.

3.2.5. Cultural

The lack of social cues and the option for anonymity in social media invoke higher aggressiveness than in offline space (Chou et al., 2014; Tynes et al., 2013). However, the line between online and offline space is unclear. The studies show that the screen time of social media was not associated with stigmatized beliefs, and the user's pre-existing stereotype is stronger predictors of stigmatization than their social learning taken place in media use (Cho et al., 2021; Tynes et al., 2013). Thus, the cause of hateful discourse such as body shaming and racist hate speech in social media platforms should include the cultural explanation. Understanding the cultural cause of how social media interactions and discourses are formed is

critical in addressing social media mental health issues, since even if it is not in social media company's direct control, it contributes to shaping the user experience.

4. Social Media Design for User Mental Health

4.1. Significance

The examination of what causes the harm on user mental health at the platform level demonstrates how social media company's impact on users is affected by the economic, legal and political motivations and the scope of technology advancement and the culture today. This shows that the cause of the user well-being problem in social media is more extensive than the platform's alone.

The Center for Humane Technology illustrates *Framework of Interventions* that exhibit the systematic solutions for more ethical technology. Its model includes the potential solutions to be Platform Changes, the change in user experience of the platform, Internal Governance, the change in company's internal operations by the internal decision makers, and Business Model, the change in profit structure of the firm, External Regulation, the change in company's operations by the external lawmakers, Economic Goal, the change in the definition of economic success and failure, and Culture and Paradigm, the change in the value system of the society.

From Platform Changes to the shift in Culture and Paradigm, though the difficulty of execution increases, its impact of change increases as well. External Regulation, Economic Goal, and Culture and Paradigm are more distant from the social media companies' control and generally more difficult to be transformed. The leverages social media companies can affect such as Platform Changes, Internal Governance, and Business Model is relatively easier to be executed for the social impact.

Moreover, a multi-level perspective of social transition that VSD is based on views the transition is made as the effect of interactions among social groups in the system (Geels & Kemp, 2007). Hence, technology has ethical significance as it shapes the context of how users act, perceive, and interact with their environment (Verbeek, 2009). The interpretation of social transition in the multi-level perspective informs that design choice of engineers matters because it affects the direction of social transition.

The effort of platforms in resolving user mental health issues matters because the difficulty of executing is relatively lower and its impact is meaningful. As mentioned earlier, platforms' design affect the user experience and thus, their well-being. Thus, while the progress in legislation, economy, and culture is made very gradually, platforms can effectively make direct changes that advocate user's well-being and affect the social discourse on conscious development of social media. And since the VSD approach is effective in identifying stakeholders, legitimate values, and value tensions based on qualitative and quantitative investigations, its effectiveness in guiding technology developers' choices, thus, impacts social transition.

4.2. Case Studies: Would VSD be Effective in Social Media Design for User Mental Health?

The aim of this evaluation is to assess how effective VSD can be for social media companies to identify stakeholders and values and design implementations for user mental health. As there was no accessible source on what evidence social media companies base their choices of designs on, I will look at two case studies, Instagram's option for hiding "Like" number and Facebook's incorporation of user well-being metrics to their algorithm, and analyze their goal, their outcome, and their effectiveness from a VSD perspective. Their goal would be

identified with the stakeholders, values, and value tensions, and their effectiveness would be measured with how much their outcomes are aligned to their goals. After the analysis, I will look at how the success or the limitations of the design can be or not be supported by VSD.

4.2.1. Case Study 1: Instagram's Option for Hiding "Like" Number

4.2.1.1. The Goal: Stakeholders and Values

"Like" is a key engagement activity in Instagram that the majority of young users use (Prichard et al., 2021). While "Like" feature can positively impact user well-being by reinforcing positive self-presentation via validation, their dependency on the feature for self-validation negatively impacts user mental health (Dumas et al., 2017). Jackson and Luchner found that self-critical users are invested more in the number of "Likes" they desire to receive, and Hart et al. suggest that users with vulnerable narcissism were negatively impacted by the "Like" feature due to the fear of being judged by others negatively (Hart et al., 2017; Jackson & Luchner, 2018). Also, the social comparison induced by comparing the number of "Likes" affected users' body image concerns (Mingoia et al., 2017). These findings are consistent with how user's prior mental health affects and is reinforced by their social media experience.

In July 2019, in response to the problem of "Likes" on user mental health, Instagram started testing users to see how hiding the number of likes impacts their well-being in Japan, Australia, Brazil, Canada, Ireland, Italy, and New Zealand (Prichard et al., 2021). As Instagram CEO Adam Mosseri announced expanding the test to the US, Mosseri claimed "We will make decisions that hurt the business if they help people's well-being and health" (So, 2019).

The mentioned stakeholders are the company and the user, and the values that are conflicting are the company's profit interest and the user's well-being and health. Mosseri also claimed that the test's intention is not to incentivize certain user engagement but to create "a less

pressurized environment where people feel comfortable expressing themselves" as they "hear people worry about how many like counts they get" (Mac & Nguyen, 2019). In other words, with the identified stakeholders and the values, Instagram prioritized the user's mental health over the company's profit and tested hiding "Like" counts to support the user's mental health.

4.2.1.2. The Outcome

The visibility of "Likes" was associated with a negative impact on user well-being such as being upset, hostile, ashamed, nervous, or afraid, and had a higher correlation when the number of "Likes" was higher (Wallace & Buil, 2021). And no evidence was found to support that the visibility of "Likes" improves user mental health (Wallace & Buil, 2021).

However, regardless of the visibility of "Likes," users felt lonely and still impacted users' well-being (Wallace & Buil, 2021). As hiding "Likes" number option did not eradicate the "Like" feature, receiving less "Likes" that desired still resulted in greater loneliness, and receiving more "Likes" than desired resulted in greater dissatisfaction and vicious Like-seeking behavior (Wallace & Buil, 2021).

Moreover, after hiding the number of "Likes," despite the concern from social media influencers or brands that there will be a downward trend in liking engagement with influencer posts, no change was observed in people's liking behavior, which shows that popular posts stayed popular even after the implementation.

Thus, from this investigation, it was found that hiding the number of "Likes" addresses the negative impact on user mental health that is compounded with the impact the number of "Likes" has on user mental health. However, it does not reduce the mental health issue triggered by Like-seeking behavior and social comparison as "Likes" and its meaning still exist and does not change the Instagram discourse on what receives more "Likes" than others.

And in May 2021, Instagram announced an option to hide public "Like" counts of both user's own posts and others' (Instagram, 2021). In that announcement, Instagram stated "We tested hiding like counts to see if it might depressurize people's experience on Instagram. What we heard from people and experts was that not seeing like counts was beneficial for some, and annoying to others, particularly because people use like counts to get a sense for what's trending or popular, so we're giving you the choice" (Instagram, 2021). Through the empirical investigation, Instagram confirmed the positive impact of hiding the number of "Likes" but also identified a new value, autonomy, and prioritized its importance as well in user experience.

4.2.1.3. The Effectiveness of the Design

The goal of hiding the "Like" count feature was to protect user well-being and health by providing them a freer space from social comparison. Though the investigation suggests that hiding the visibility of the number of "Likes" positively impacts user mental health, it does not address the remaining negative impact of "Like" count on user mental health. Instagram's final decision to make the feature optional for users addresses the new value, autonomy, identified from the empirical investigation but still does not address the remaining impact of "Likes" on user mental health, which was the original goal of the feature. Therefore, there was no action after the questions why users are still negatively impacted by "Like" counts and why their liking behavior has not changed even when they are not visible to the public by Instagram.

4.2.1.4. Implementation of VSD

"It turned out that it didn't actually change nearly as much about … how people felt, or how much they used the experience as we thought it would. … But it did end up being pretty polarizing. Some people really liked it, and some people really didn't. … It ended up being that

the clearest path forward was something that we already believe in, which is giving people choice," said Mosseri (Newton, 2021).

Instagram's implementation of hiding "Like" counts showed that the negative impact of "Like" on user mental health was not solely reduced by making the number of "Likes" invisible (Newton, 2021; Wallace & Buil, 2021). Instagram's response to such a finding that has made hiding "Like" counts optional for users for their own distinct, satisfying experience leaves a question whether or not user autonomy fully addresses the user mental health issue affected by "Likes" via this feature.

After identifying the stakeholders and their values at the conceptual investigation, VSD's empirical investigation highlights what values stakeholders prioritize among various value tensions when they actually use the technology. And to the unanswered question of why there was no change of impact of "Likes" on users, hiding "Like" counts lacks on addressing what users seek for through "Like" counts.

When it comes to the value tensions of users in Like-seeking behavior, despite the lower self-esteem Like-seeking behavior contributes to, users are observed to compromise their social needs of genuine connection with peers with their social needs of self-validation and a sense of peer belonging via receiving a higher number of "Likes" (Dumas et al., 2020). Thus, the count of "Likes" still matters to the users, especially to the ones who have been invested in Like-seeking, even when the number is not visible as its meaning of affirmation and peer belonging, which fosters deceptive Like-seeking behavior, exists.

Instagram's test of hiding the number of "Likes" has recognized the value tension between company's profit interest and user mental health but it has not identified the value tension within the user when engaging with "Like" features, which resulted in a minimal change

after the implementation of the test. This case study manifests that to successfully address user mental health issues triggered or exacerbated by social media requires more extensive understanding of stakeholder's values. Therefore, VSD's role of recognizing values and value tensions even within one stakeholder would be helpful for social media design for user mental health with more precise approach to the causes of the issue.

4.2.2. Case Study 2: Facebook's Incorporation of Meaningful Social Interaction (MSI) metric to Its News Feed Algorithm

4.2.2.1. The Goal: Stakeholders and Values

Studies show that while user's active social interaction in Facebook is correlated with improved user well-being, user's passive use of Facebook, a passive consumption of others' contents rather than an active interaction with others, exacerbates user mental health, including loneliness, social anxiety, and depression, due to an increased frequency of social comparison and a lack of strong social bonding (Burke et al., 2010; Shaw et al., 2015; Tosun & Kaşdarma, 2020). Tosun and Kaşdarma (Tosun & Kaşdarma, 2020) further specify this correlation that the effect of passive usage on user mental health depends on if the social comparison was targeted to their acquaintances or to their close friends as the frequency of depression is higher in the former case. Thus, Facebook user's mental health is affected by how they use Facebook and who they interact with.

The leaked Facebook's 2018 internal research (Nix & Wagner, 2022) also acknowledged that Facebook use made users more lonely than other activities such as using Instagram/YouTube/Twitter, spending time with friends and family, playing video games, using dating apps, and watching TV (Nix & Wagner, 2022). And the research also showed while 41% of their users felt less lonely after Facebook use, 6% of them were more lonely and 42% of them

were more or less lonely (Nix & Wagner, 2022). Hence, based on the research, Facebook concluded that Facebook use can be beneficial for user well-being if it is used for "sharing messages, posts and comments with close friends and reminiscing about past interactions" (Ginsberg & Burke, 2017; Seetharaman, 2017).

In 2018, Facebook CEO, Mark Zukerberg, uploaded a Facebook post that the company has reoriented its goal from recommending users relevant contents that might interest them to supporting users to build "meaningful social interactions" (Zuckerberg, 2018). With this post, Facebook announced its redesigned new feed algorithm that prioritizes the posts' ranking that initiate meaningful interactions (Mosseri, 2018). Defined by Meta, meaningful social interactions are "interactions that people believe enhance their lives, the lives of their interaction partners, or their relationships, with emotional, informational, or tangible impact" (Litt et al., 2020, p.9). How they measure the meaningfulness of the post was via a point system that categorized the types of engagement activity and the strength of the bond users are interacted with (Metz, 2021). According to the leaked Facebook document (Hagey & Horwitz, 2021), titled "The story of deriving the Meaningful Social Interactions metric weights (UX Research & Data Science)," each "like" is considered 1 point, a reaction emoji 5 points, and comments that have at least 5 unique tokens 30 points (Hagey & Horwitz, 2021). Then, this engagement activity point would be multiplied by the factor that reflects the closeness of the relationship such as by 0.3 if they are strangers and by 0.5 if they belong to the same Facebook group but are not a friend (Hagey & Horwitz, 2021).

In the post (Zuckerberg, 2018), Zukerberg explained that the change was based on the findings that demonstrated a positive correlation between social media user's well-being and their use for interacting with the people they care about (Zuckerberg, 2018). A relevant research

in 2016 (Burke & Kraut, 2016) also investigated how the type of communication and the tie strength with people, who social media users interact with, affect users' well-being. It demonstrates that while personalized, composed communication from strong ties improve users' well-being, that form of communication from weak ties or a broadcast or one-click communication regardless of the bond strength do not affect users' well-being (Burke & Kraut, 2016).

The stakeholders of this implementation are users, Facebook, and businesses that use the Facebook algorithm for advertising. And the values that are at stake are user well-being, Facebook's interest in profit and success in business, and those businesses' profit interest. In the post (Zuckerberg, 2018), Zukerberg mentioned Facebook's potential profit loss as a tradeoff of this implementation for user well-being. However, the 2018 leaked internal document (Hagey & Horwitz, 2021), titled "Key FAST Goal Metrics," suggested that there was also a Facebook's business incentive to encourage users' interactions in comments, likes, and reshares by making engageable posts more visible because there had been a general decline in user engagement through 2017 to 2018 (Hagey & Horwitz, 2021). Hence, Facebook's short-term potential profit loss from reduced user screen time was for Facebook's recovery in user engagement and ultimately, more sustainable business. And in case of the businesses that harness Facebook to reach potential consumers, since MSI metric would rank the posts from weak ties that generate less user interaction lower, businesses' profit interest is potentially compromised for user mental health. Therefore, Facebook's goal of this design change is to support user mental health and to overcome decline in its user engagement rate.

4.2.2.2. The Outcome

Unlike Facebook's expectation on potential trade-off on user engagement after applying MSI metric, Facebook succeeded in increasing the user engagement by 50% in 2019 (Grothaus, 2019). However, prioritizing the posts that attracted more comments, reactions, likes, and reshares did not serve user well-being but rather it resulted in increased prevalence of angrier contents (Grothaus, 2019; Hagey & Horwitz, 2021). For the businesses, as expected by Facebook, experienced a decline in their user outreach in the first half of 2018 after the implementation (Hagey & Horwitz, 2021). Therefore, MSI metric incentivized businesses to produce harmful contents that increase user-engagement, which contain misinformation, toxicity, and violence (Hagey & Horwitz, 2021). Facebook internal research, titled "Does Facebook reward outrage? Posts that generate negative comments get more clicks," also noticed that the content publishers had a financial incentive to capitalize on the negativity to generate more user outreach (Metz, 2021).

After launching the addition of MSI metric, Facebook internal researchers have continuously revised the algorithm to make MSI more effective to its goal of enhancing user mental health (Hagey & Horwitz, 2021; Horwitz, n.d.). To address the rapid proliferation of misinformation and violence caused by MSI metric, in 2019, it zeroed the worth of the reshare activity when it is done by a viewer who is not a friend or a follower of an original poster, which turned out to be effective in cutting the spread of misinformation in its early testing applied to civic and health contents and to some countries such as Ethiopia and Myanmar (Hagey & Horwitz, 2021). And in 2020, to target the financial incentive of businesses to produce negative contents, the angry reaction was recalculated as a zero value (Metz, 2021). And from 2019, to create a more curated space for user's meaningful time spent on Facebook, Facebook added a

survey feature that asks users whose and what posts they want to see more (Sethuraman et al., 2019).

4.2.2.3. The Effectiveness of the Design

The goal of applying MSI metric in 2018 was only partially met. Facebook's goal to reincrease user engagement was achieved through MSI metric, but its goal to enhance user well-being in Facebook use failed with proliferation of more divisive and toxic contents. However, it is critical to note that this design is a work in progress and the design is continuously revised by Facebook internal researchers with the aim of supporting user and community well-being (i.e. zeroing the worth of angry reaction to disincentivize businesses' toxic content production for more user outreach) and giving users more control of their Facebook experience (i.e. survey). Therefore, despite its failure, it has become an explicit product Facebook researchers can address and work on for user well-being.

4.2.2.4. Implementation of VSD

VSD's aim is to not find the perfect method that would deliver the intended values with no side-effects. It understands the complexity of the cause of the problem and stresses the commitment to update the design to better align with the value. In this design, though the initial point system that MSI metric was based on user surveys on what interactions they find meaningful in Facebook, its outcome reveals the complexity of understanding how to promote user well-being. And the following course of updates applied to MSI metric shows how the commitment to update the design for the aimed value lets the failures and unintended effects become a guidance to better facilitate healthy user engagement, which reinforces VSD's point.

Moreover, MSI metric's implementation manifested the bottleneck of the platform's development for user well-being, which are Facebook's internal value tension between user

well-being and profit and its organizational structure of decision-making (Hagey & Horwitz, 2021; Horwitz, n.d.; Kang & Frenkel, n.d.; Oremus, 2021). In 2019, when Anna Stepanov, a director of Facebook product management, suggested applying the changes in MSI metric to solve the wide spread of harmful contents to Zukerberg, Zukerberg claimed that the change would be pushed only if it does not harm the company's profit (Hagey & Horwitz, 2021). The reports and interviews of Facebook employees show that the differing priority between those of Facebook integrity team and those of executives, who make the final call for the platform design, make Facebook's development for user well-being more difficult (Hagey & Horwitz, 2021; Horwitz, n.d.; Kang & Frenkel, n.d.; Oremus, 2021). However, while VSD emphasizes commitment to a design, it does not incentivize all developers involved in the design to commit to a certain value. Even though Facebook researchers internally find a potential solution for the harm they have caused via VSD approach of long commitment of investigation and updates, if Facebook managerial structure becomes the barrier, VSD cannot be fully executed and the design for user mental health would be slowed down. Hence, this demonstrates that for user mental health and well-being, it needs a change not only in platform design but also in internal governance, external regulation, business model, economic model and culture.

5. Conclusion

As technology rapidly grew, it raised an ethical concern on its impact on society such as violated privacy, under-informed consent, and harmed user well-being. From this discourse of information systems design and human-computer interaction, based on the understanding of co-evolution of technology and the society's values, Value Sensitive Design was modeled to guide a conscious design via committed investigations and recognitions of all stakeholders and value tensions.

Social media is a platform where anyone can produce content and interact with each other. And though the causal relationship between its usage and user mental health is ambiguous and complex, its platform design choices affect user experience and thus, the well-being of user mental health. Its harm on user mental health is caused by economic, legal, political, technical, and cultural factors, where the first three reflect company's interest in profit and the last two show the hindrance in technological and societal progress for improving user mental health.

Based on the *Framework of Interventions* by The Center for Humane Technology, platform level effort for resolving user mental health issues is critical as its impact is significant and the difficulty of executing is relatively lower than other legislative, economic, societal efforts. VSD's multi-lifespan viewpoint also supports platform level effort as it facilitates social progress. Hence, at a platform, where myriads of value tensions and stakeholders thrive, to address the social media's user mental health issue, this paper started with a question: would VSD be effective in social media for user mental health?

The paper has done two case studies, one on Instagram's hiding "Like" feature and another on Facebook's "Meaningful Social Interaction" metric, to study how the design well-addressed their stakeholders, value tensions, and its original goal and how VSD can or cannot complement the design. According to the result of case studies, VSD approach for social media user mental health is highly recommended as its guidance on rigorous conceptual, empirical, and technical investigation helps comprehend value tensions and stakeholders for a more precise approach to solving user mental health issues. Also, VSD's multi-lifespan stance to a design resonates with the realistic outcome of design failures and the successes from a long-term commitment and updates to a design. However, as mentioned above, social media's user mental health problem is substantially caused by the company's interest in profit. If the

company's business goal and internal governance do not change, the internal platform design change effort for user well-being, even with VSD, has limited effectiveness. Hence, such a limitation of VSD underlines VSD's multi-life span viewpoint and *Framework of Interventions* as fundamentally business, legislative, economic, societal change should be made for sustainable and effective user well-being.

5.1. Limitation of the Paper and What's Next

While co-evolution of technology and society perspective is significant in VSD, the paper's case study on the design's outcome does not contain a comprehensive understanding of how the design influenced other societal efforts such as legislative or cultural change for user well-being. Also, while the Facebook case study's analysis on outcome discusses the diminished user well-being with increased harmful contents, it does not have a direct source on how MSI metric affected user loneliness, social anxiety, and depression. Moreover, as it did not have direct access to how each social media company's design workflow looked, the evaluation on how VSD could complement their design approach is not complete.

Future research can examine the co-evolution of social media change and societal change for user mental health to have a deeper understanding of the platform design's significance. Also, with a direct source such as interviews of social media company's internal researchers, it can more accurately evaluate if VSD can or cannot complement social media design approach for user mental health.

Berry, N., Lobban, F., Belousov, M., Emsley, R., Nenadic, G., & Bucci, S. (2017). #WhyWeTweetMH: Understanding Why People Use Twitter to Discuss Mental Health Problems. *Journal of Medical Internet Research*, *19*(4), e6173. https://doi.org/10.2196/jmir.6173

Bhargava, V. R., & Velasquez, M. (2021). Ethics of the Attention Economy: The Problem of Social Media Addiction. *Business Ethics Quarterly*, *31*(3), 321–359. https://doi.org/10.1017/beq.2020.32

Bødker, S., & Pekkola, S. (2010). *A short review to the past and present of participatory design*. 4.

Brannon, V. C. (2019). *Free Speech and the Regulation of Social Media Content*. 47.

Briggs, P., & Thomas, L. (2015). An Inclusive, Value Sensitive Design Perspective on Future Identity Technologies. *ACM Transactions on Computer-Human Interaction*, *22*(5), 1–28. https://doi.org/10.1145/2778972

Burke, M., & Kraut, R. E. (2016). The Relationship Between Facebook Use and Well-Being Depends on Communication Type and Tie Strength. *Journal of Computer-Mediated Communication*, *21*(4), 265–281. https://doi.org/10.1111/jcc4.12162

Burke, M., Marlow, C., & Lento, M. (2010). Social network activity and social well-being. In *Conference on Human Factors in Computing Systems—Proceedings* (Vol. 3, p. 1912). https://doi.org/10.1145/1753326.1753613

Cameron, D., Wodinsky, S., & DeGeurin, M. (2022, April 18). *Read the Facebook Papers for Yourself*. Gizmodo. https://gizmodo.com/facebook-papers-how-to-read-1848702919

Cho, H., Li, W., Cannon, J., Lopez, R., & Song, C. (Chuck). (2021). Testing three explanations for stigmatization of people of Asian descent during COVID-19: Maladaptive coping,

biased media use, or racial prejudice? *Ethnicity & Health*, *26*(1), 94–109. https://doi.org/10.1080/13557858.2020.1830035

Chou, W. S., Prestin, A., & Kunath, S. (2014). Obesity in social media: A mixed methods analysis. *Translational Behavioral Medicine*, *4*(3), 314–323. https://doi.org/10.1007/s13142-014-0256-1

Cohen, R., & Blaszczynski, A. (2015). Comparative effects of Facebook and conventional media on body image dissatisfaction. *Journal of Eating Disorders*, *3*(1), 23. https://doi.org/10.1186/s40337-015-0061-3

Datta, A., Buchegger, S., Hung, L., Strufe, T., & Rzadca, K. (2010). *Decentralized Online Social Networks* (pp. 349–378). https://doi.org/10.1007/978-1-4419-7142-5_17

Dumas, T. M., Maxwell-Smith, M. A., Tremblay, P. F., Litt, D. M., & Ellis, W. (2020). Gaining likes, but at what cost? Longitudinal relations between young adults' deceptive like-seeking on instagram, peer belonging and self-esteem. *Computers in Human Behavior*, *112*, 106467. https://doi.org/10.1016/j.chb.2020.106467

Dumas, T. M., Maxwell-Smith, M., Davis, J. P., & Giulietti, P. A. (2017). Lying or longing for likes? Narcissism, peer belonging, loneliness and normative versus deceptive like-seeking on Instagram in emerging adulthood. *Computers in Human Behavior*, *71*, 1–10. https://doi.org/10.1016/j.chb.2017.01.037

Facebook in the News by Machine Learning in Action. (n.d.). Anchor. Retrieved December 5, 2022, from https://anchor.fm/lily-adelstein/embed/episodes/ML-in-the-News-Facebook-e18c989

Faelens, L., Hoorelbeke, K., Cambier, R., van Put, J., Van de Putte, E., De Raedt, R., & Koster, E. H. W. (2021). The relationship between Instagram use and indicators of mental health:

A systematic review. *Computers in Human Behavior Reports*, *4*, 100121. https://doi.org/10.1016/j.chbr.2021.100121

Frank, R. H. (2021, February 11). The Economic Case for Regulating Social Media. *The New York Times*. https://www.nytimes.com/2021/02/11/business/social-media-facebook-regulation.html

Friedman, B., & Hendry, D. G. (2019). *Value Sensitive Design: Shaping Technology with Moral Imagination*. MIT Press.

Friedman, B., Kahn, P. H., & Borning, A. (2002). *Value Sensitive Design: Theory and Methods*. 8.

Friedman, B., Kahn, P. H., & Borning, A. (2013). *Value Sensitive Design and Information Systems*. 27.

Geels, F. W., & Kemp, R. (2007). Dynamics in socio-technical systems: Typology of change processes and contrasting case studies. *Technology in Society*, *29*(4), 441–455. https://doi.org/10.1016/j.techsoc.2007.08.009

George, M. (2019). The Importance of Social Media Content for Teens' Risks for Self-harm. *Journal of Adolescent Health*, *65*(1), 9–10. https://doi.org/10.1016/j.jadohealth.2019.04.022

Gillespie, T. (2010). The politics of 'platforms.' *New Media & Society*, *12*, 347–364. https://doi.org/10.1177/1461444809342738

Ginsberg, D., & Burke, M. (2017, December 15). Hard Questions: Is Spending Time on Social Media Bad for Us? *Meta*. https://about.fb.com/news/2017/12/hard-questions-is-spending-time-on-social-media-bad-for-us/

Griffiths, M. D. (2018). Adolescent social networking: How do social media operators facilitate habitual use? *Education and Health*, *36*(3), Article 3.

Grothaus, M. (2019, March 18). *Facebook's "meaningful interaction" news feed change is a failure*. https://www.fastcompany.com/90321304/facebooks-meaningful-interaction-newsfeed-change-is-a-failure

Hagey, K., & Horwitz, J. (2021, September 15). Facebook Tried to Make Its Platform a Healthier Place. It Got Angrier Instead. *Wall Street Journal*. https://www.wsj.com/articles/facebook-algorithm-change-zuckerberg-11631654215

Hart, W., Adams, J., Burton, K. A., & Tortoriello, G. K. (2017). Narcissism and self-presentation: Profiling grandiose and vulnerable Narcissists' self-presentation tactic use. *Personality and Individual Differences*, *104*, 48–57. https://doi.org/10.1016/j.paid.2016.06.062

Horwitz, J. (n.d.). *Facebook Knew Calls for Violence Plagued 'Groups,' Now Plans Overhaul*. WSJ. Retrieved December 6, 2022, from https://www.wsj.com/articles/facebook-knew-calls-for-violence-plagued-groups-now-plans-overhaul-11612131374

Instagram. (2021, May 26). *Giving People More Control on Instagram and Facebook*. https://about.instagram.com/blog/announcements/giving-people-more-control

Jackson, C. A., & Luchner, A. F. (2018). Self-presentation mediates the relationship between Self-criticism and emotional response to Instagram feedback. *Personality and Individual Differences*, *133*, 1–6. https://doi.org/10.1016/j.paid.2017.04.052

Kang, C., & Frenkel, S. (n.d.). *An Ugly Truth: Inside Facebook's Battle for Domination*.

Retrieved December 7, 2022, from https://read.amazon.com/kp/embed?preview=newtab&linkCode=kpe&reshareChannel=system&reshareId=98XVRZS2AT673BZA820W&ref_=cm_sw_r_kb_dp_QQZ5EHWRZD7Z5HNMG1N3&asin=B07YSHD2C5&tag=thewaspos09-20&amazonDeviceType=A2CLFWBIMVSE9N

Kapoor, K. K., Tamilmani, K., Rana, N. P., Patil, P., Dwivedi, Y. K., & Nerur, S. (2018). Advances in Social Media Research: Past, Present and Future. *Information Systems Frontiers*, *20*(3), 531–558. https://doi.org/10.1007/s10796-017-9810-y

Keles, B., McCrae, N., & Grealish, A. (2020). A systematic review: The influence of social media on depression, anxiety and psychological distress in adolescents. *International Journal of Adolescence and Youth*, *25*(1), 79–93. https://doi.org/10.1080/02673843.2019.1590851

Lewis, P. (2017, October 6). "Our minds can be hijacked": The tech insiders who fear a smartphone dystopia. *The Guardian*. https://www.theguardian.com/technology/2017/oct/05/smartphone-addiction-silicon-valley-dystopia

Litt, E., Zhao, S., Kraut, R., & Burke, M. (2020, August 11). *What Are Meaningful Social Interactions in Today's Media Landscape? A Cross-Cultural Survey - Meta Research*. Meta Research. https://research.facebook.com/publications/what-are-meaningful-social-interactions-in-todays-media-landscape-a-cross-cultural-survey/

Mac, R., & Nguyen, N. (2019, April 30). *Instagram Is Hiding Like Counts From Followers In A Test To See If You'll Feel Less Horrible*. BuzzFeed News.

https://www.buzzfeednews.com/article/ryanmac/instagram-hide-like-counts-in-test

Metz, R. (2021, October 27). *Likes, anger emojis and RSVPs: The math behind Facebook's News Feed — and how it backfired | CNN Business*. CNN. https://www.cnn.com/2021/10/27/tech/facebook-papers-meaningful-social-interaction-news-feed-math/index.html

Mingoia, J., Hutchinson, A. D., Wilson, C., & Gleaves, D. H. (2017). The Relationship between Social Networking Site Use and the Internalization of a Thin Ideal in Females: A Meta-Analytic Review. *Frontiers in Psychology*, *8*, 1351. https://doi.org/10.3389/fpsyg.2017.01351

Mosseri, A. (2018, January 11). Bringing People Closer Together. *Meta*. https://about.fb.com/news/2018/01/news-feed-fyi-bringing-people-closer-together/

Napoli, P. M. (2019). *Social Media and the Public Interest: Media Regulation in the Disinformation Age*. Columbia University Press.

Newton, C. (2021, May 27). *What Instagram really learned from hiding like counts*. The Verge. https://www.theverge.com/2021/5/27/22456206/instagram-hiding-likes-experiment-results-platformer

Nix, N., & Wagner, K. (2022, March 2). Facebook Researchers Find Its Apps Can Make Us Lonelier. *Bloomberg.Com*. https://www.bloomberg.com/news/features/2022-03-02/facebook-researchers-find-its-apps-can-make-us-lonelier

Oremus, W. (2021, September 16). Analysis | Facebook keeps researching its own harms—And burying the findings. *Washington Post*. https://www.washingtonpost.com/technology/2021/09/16/facebook-files-internal-research

-harms/

Padín, P. F., González-Rodríguez, R., Verde-Diego, C., & Vázquez-Pérez, R. (2021). Social media and eating disorder psychopathology: A systematic review. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, *15*(3), Article 3. https://doi.org/10.5817/CP2021-3-6

Prichard, I., O'Toole, S., Wu, Y., Harford, J., & Tiggemann, M. (2021). No likes, no problem? Users' reactions to the removal of Instagram number of likes on other people's posts and links to body image. *Body Image*, *38*, 72–79. https://doi.org/10.1016/j.bodyim.2021.03.007

Schechner, S., Horwitz, J., & Glazer, E. (2021, September 17). How Facebook Hobbled Mark Zuckerberg's Bid to Get America Vaccinated. *Wall Street Journal*. https://www.wsj.com/articles/facebook-mark-zuckerberg-vaccinated-11631880296

Seetharaman, D. (2017, December 15). *Facebook Acknowledges Some Social-Media Use Is Harmful*. WSJ. https://www.wsj.com/articles/facebook-acknowledges-some-social-media-use-is-harmful-1513379765

Sethuraman, R., Vallmitjana, J., & Levin, J. (2019, May 16). Using Surveys to Make News Feed More Personal. *Meta*. https://about.fb.com/news/2019/05/more-personalized-experiences/

Shaw, A. M., Timpano, K. R., Tran, T. B., & Joormann, J. (2015). Correlates of Facebook usage patterns: The relationship between passive Facebook use, social anxiety symptoms, and brooding. *Computers in Human Behavior*, *48*, 575–580. https://doi.org/10.1016/j.chb.2015.02.003

So, A. (2019, November 9). *Instagram Will Test Hiding "Likes" in the US Starting Next Week |*

WIRED. https://www.wired.com/story/instagram-hiding-likes-adam-mosseri-tracee-ellis-ross-wired25/

TIER0_RANK_RO_0920. (n.d.). Retrieved December 8, 2022, from https://www.documentcloud.org/documents/21601629-tier0_rank_ro_0920

Tosun, L. P., & Kaşdarma, E. (2020). Passive Facebook Use and Depression. *Journal of Media Psychology*, *32*(4), 165–175. https://doi.org/10.1027/1864-1105/a000269

Tynes, B. M., Rose, C. A., & Markoe, S. L. (2013). Extending campus life to the Internet: Social media, discrimination, and perceptions of racial climate. *Journal of Diversity in Higher Education*, *6*, 102–114. https://doi.org/10.1037/a0033267

Verbeek, P.-P. (2009). The Moral Relevance of Technological Artifacts. In P. Sollie & M. Düwell (Eds.), *Evaluating New Technologies: Methodological Problems for the Ethical Assessment of Technology Developments.* (pp. 63–77). Springer Netherlands. https://doi.org/10.1007/978-90-481-2229-5_6

Verduyn, P., Lee, D. S., Park, J., Shablack, H., Orvell, A., Bayer, J., Ybarra, O., Jonides, J., & Kross, E. (2015). Passive Facebook usage undermines affective well-being: Experimental and longitudinal evidence. *Journal of Experimental Psychology: General*, *144*(2), 480–488. https://doi.org/10.1037/xge0000057

Wallace, E., & Buil, I. (2021). Hiding Instagram Likes: Effects on negative affect and loneliness. *Personality and Individual Differences*, *170*, 110509. https://doi.org/10.1016/j.paid.2020.110509

Wiener, N. (1988). *The Human Use Of Human Beings: Cybernetics And Society*. Hachette Books.

Zuckerberg, M. (2018, January 11).

https://www.facebook.com/zuck/posts/pfbid0KmmxDGszdBajmdCCYWDHTWtfw1nNxLZdCzTTHAHa9K6Xb1AewJzw7mV71vL2SNx3l