



Governing principles: Articulating values in social media platform policies

new media & society

1–20

© The Author(s) 2023



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/14614448231156580

journals.sagepub.com/home/nms

Rebecca Scharlach* , Blake Hallinan* 
and Limor Shifman 

The Hebrew University of Jerusalem, Israel

Abstract

As sites where social media corporations profess their commitment to principles like community and free speech, policy documents function as boundary objects that navigate diverse audiences, purposes, and interests. This article compares the discourse of values in the Privacy Policies, Terms of Service, and Community Guidelines of five major platforms (Facebook, Instagram, YouTube, Twitter, and TikTok). Through a mixed-methods analysis, we identified frequently mentioned value terms and five overarching principles consistent across platforms: *expression*, *community*, *safety*, *choice*, and *improvement*. However, platforms limit their burden to execute these values by selectively assigning responsibility for their enactment, often unloading such responsibility onto users. Moreover, while each of the core values may potentially serve the public good, they can also promote narrow corporate goals. This dual framing allows platforms to strategically reinterpret values to suit their own interests.

Keywords

boundary objects, Facebook, Instagram, platform policies, TikTok, Twitter, values, YouTube

As multi-billion dollar companies, social media platforms strive to create value; as spaces where billions of people communicate, they seek to promote particular values. The public significance of platforms stems, to a large extent, from the intersection of the

*Rebecca Scharlach and Blake Hallinan contributed equally to this study.

Corresponding author:

Rebecca Scharlach, Department of Communication and Journalism, The Hebrew University of Jerusalem, Mount Scopus, 91905 Jerusalem, Israel.

Email: rebecca.scharlach@mail.huji.ac.il

economic and moral meanings of “value,” with social media simultaneously driving global markets and shaping what people find important. While the economic valuation of digital platforms has been the subject of significant academic interest (e.g. Birch et al., 2021; Van Dijck et al., 2018), the relationship between platforms and values remains ambiguous even in the face of growing public concern (Hallinan et al., 2022). Tracing the construction of values by and through platforms has thus become an important but challenging imperative.

Industry professionals increasingly recognize that platforms shape ethical commitments (DeNardis and Hackl, 2015). Indeed, tech executives do not shy away from publicly invoking values. For instance, Facebook’s mission statement includes the “core values” of the company: be bold, focus on impact, move fast, be open, and build social value (Facebook, 2015). In a similar vein, Adam Mosseri, CEO of Instagram, tweeted that “We’re not neutral. No platform is neutral, we all have values and those values influence the decisions we make” (Mosseri, 2021). Although companies tend to frame “positive” values as fundamental to their operations, controversies over privacy, political bias, and the algorithmic amplification of extreme content have generated fierce debates over the actual values platforms promote.

Alongside growing public discourse, scholars across disciplines have flagged the importance of studying the relationship between values and social media platforms (Gillespie et al., 2020; Leurs and Zimmer, 2017). Despite this interdisciplinary interest, platform values remain slippery objects of analysis due to their conceptual complexity and the diverse stakeholders involved. Given these challenges, we suggest that policy documents serve as productive sites for exploring platform values. As public-facing and idealized accounts of how various actors should behave, platform policies reflect and facilitate the negotiation of values. Although research has investigated how users understand policy documents (Quinn et al., 2019) and the role of policies within the broader ecosystem of platform governance (Gorwa, 2019), the articulation of values in policies has evaded systematic analysis.

Addressing this gap, we present the first overview of the values invoked in the policies of five major social media platforms: Facebook, Instagram, YouTube, Twitter, and TikTok. We start by surveying the literature on the multi-stakeholder construction of values on social media platforms, positioning the policy documents of these corporations as “boundary objects” (Star and Griesemer, 1989) that cater to different parties and interests. Next, we outline our inductive approach to analyzing the discourse of values in such documents. Our findings show that policies consistently express a set of core principles that serve both commercial and public interests, and assign significant responsibility to users to uphold these values. Together, these strategies enable platforms to emphasize positive principles without jeopardizing their commercial interests.

Literature review

Platform values as a site of contestation

Platforms, broadly defined as digital infrastructures that “host, organize and circulate user’s shared content or social exchanges” (Gillespie, 2017: 417), structure possibilities for expression and interaction. Although social media companies have made use of the

“semantic richness” (Van Dijck, 2013: 349) of the term platform to offer “a comforting sense of technical neutrality and progressive openness” (p. 360) and disavow potential responsibilities, a number of high-profile controversies, as well as statements by the owners of these platforms, have made the myth of neutrality less tenable. Over the past decade, scholars have highlighted the importance of studying the values associated with social media platforms (e.g. Gillespie et al., 2020; Leurs and Zimmer, 2017; Van Dijck et al., 2018). However, studying platform values, broadly defined as the underlying principles governing and expressed through social media (Hallinan et al., 2022), is challenging. This difficulty stems from disagreements around the nature of values, the divergent stakeholders involved in social media, and the role of economic interests in shaping decisions and policies in such spheres.

The ambiguity surrounding platform values reflects an enduring disagreement over how to conceptualize values. Values are a constitutive part of being human and have generated wide interest and study for the past 200 years. Nevertheless, the exact definition of “values” is still debated (Joas, 1997; Martin and Lembo, 2020). This has to do, in part, with the duality of values as constructs that reside between thought and action. While values have been widely approached as abstract perceptions (e.g. Schwartz, 2012), their existence depends on dissemination through various forms of expression. As such, the construction of values is entangled with norms of communication in private and public settings (Hallinan et al., 2022). The pragmatic turn in sociology offers a productive path for studying the articulation of values, as it shifts the focus from values as stable personal attributes to values as discursive constructs that are performed and debated in specific contexts (Boltanski and Thévenot, 2006; Heinich, 2020; Lamont, 2012). In what follows, we build on Heinich’s differentiation between three understandings of value. *Value as worth* refers to the evaluation of something as significant, often in monetary terms; *value as object* refers to tangible articles or conceptual ideals that are consistently evaluated positively; and *value as principle* refers to underlying criteria that guide judgment. As detailed below, we find this distinction particularly helpful for the analysis of platform policies, as it allows us to differentiate between what platforms invoke as important (*values as objects*) and the governing logics directing evaluation (*values as principles*).

While pragmatic sociology offers a useful theoretical framework for investigating platform values, the involvement of diverse stakeholders introduces further complications. Recent theorizations posit a “governance triangle” where the responsibility for platforms is adjudicated through the interactions of states, firms, and nongovernmental organizations, each with their own interests (Gorwa, 2019). Accordingly, platforms put on “many faces” (Arun, 2022) and have to balance commercial interests and legal obligations with the mediation of the public and private communication of billions of users. Furthermore, these faces can and do come into conflict, as illustrated by the contrast between how companies like Meta pitch themselves as “family” whose platforms create “community” (Hallinan, 2021) and the challenging working conditions of commercial content moderators tasked with cleaning up the excesses of said “community” (Roberts, 2019).

This last example demonstrates that stakeholder conflicts often derive from tensions involving economic interests. While social media companies are revenue-driven, their

business models and market value depend on their ability to control user-related information. Yet the translation of personal information to economic value has never been simple or straightforward (Gandy, 2011). As a result, tech companies have constantly been engaged with what Birch et al. (2021) refer to as “assetization,” or the turning of personal information into capitalized property. Describing the various mechanisms of this transformation as “techcraft,” the authors demonstrate how companies refrain from directly monetizing personal data and instead stress the notion of the “user” as a tangible unit that can be measured through specific patterns of engagement.

The framing of people as “users” is typical to the discursive juggling of social media corporations and touches upon the fracture between public values and commercial interests (Van Dijck et al., 2018). Public value, a concept coined by Moore (1995), refers to the value that an organization contributes to the common good. In the context of social media, such values could include fairness, accountability, or transparency. Yet the polysemy of such concepts often allows corporations to conflate public value with their own economic or commercial interests, as scholars have demonstrated through the analysis of terms such as “sharing” (John, 2016), “engagement” (Hallinan et al., 2022), “connectivity” (Van Dijck, 2013), and “privacy” (Epstein et al., 2014). In what follows, we argue that such clashes are particularly pertinent to these corporations’ policy documents.

Policy documents as boundary objects

Contestation over platform values does not take place in a vacuum. People encounter the technical structure of platforms *in media res*. As Andersson Schwarz (2022) notes, “interpretations of infrastructures are almost always closely intertwined with semiotic statements ‘steering’ the initial prescriptive force of the technology in different directions” (p. 134). Accordingly, many stakeholders orient around the professed commitments of social media platforms expressed through official policy documents. While policies do not directly determine the behavior of users on the platform, they present “owner expectations governing user participation on matters ranging from acceptable content and copyrights to uses of personal information and user policy formation” (Stein, 2013: 354). In addition to offering a public-facing account of the roles and responsibilities of key actors, policies are strategic documents that aim to protect the platform from liability and respond to the broader regulatory ecosystem.

Given their interstitial position, navigating different audiences, purposes, and interests, platform policies can be understood as boundary objects. A seminal concept from science and technology studies, boundary objects are defined as “scientific objects which both inhabit several intersecting social worlds. . . and satisfy the informational requirements of each of them” (Star and Griesemer, 1989: 393). To do so, boundary objects must be flexible enough to be adapted to different needs while still retaining a core identity. Platforms clearly traverse different social worlds, given the global reach of social media, the different parties involved in creating policy documents, and the established tensions between the values of different stakeholders within an organization (Wong, 2021).

Just as boundary objects are necessarily under-specified in order to travel to different contexts (Star and Griesemer, 1989), evidence of under-specification is a major source of

criticism for platform policies. For example, Kopf (2022) shows how the “strategic vagueness” of YouTube’s policies generates uncertainty around what content is eligible for monetization, makes it difficult to contest content moderation decisions, and may even “boost creators’ (self)censorship” (p. 14). Along a similar vein, comparative platform policy analyses found that no platforms offer a concrete definition of harassment (Pater et al., 2016) or harm (DeCook et al., 2022). This vagueness simultaneously allows platforms to respond more easily to unfolding events and frames harm as a consequence of bad actors, which bounds the responsibility of the platform (DeCook et al., 2022).

Platform policies change in response to legal regimes. For example, platforms “splinter” their approach to content moderation to comply with national laws (Ahn et al., 2022) and adjust the deployment of election-related features to avoid negative publicity and regulatory scrutiny, exemplifying a situation that Barrett and Kreiss (2019) term “platform transience.” Technological developments and market conditions also drive policy changes, dramatically illustrated by the upheaval on Twitter following the company’s takeover by Elon Musk in October 2022. Such changes occasionally prompt users to contest the values of a platform, including an early protest of Yahoo’s changes to the Terms of Service (TOS) on GeoCities (Reynolds and Hallinan, 2021) and more recent responses to WhatsApp’s Privacy Policy updates in 2016 (Fiesler and Hallinan, 2018) and 2021 (Griggio et al., 2022). While such backlashes highlight users’ aspirations to participate in platform governance, involvement is not limited to highly visible protests. Localized initiatives include developing collective strategies for privacy self-management (Baruh and Popescu, 2017) or doing “anti-racist work” such as reporting and performing free content moderation of harmful content as a “labor of repair” (Nakamura, 2021).

Given the dynamic nature of platform policies, we aim to identify enduring constructs and justifications invoked in these documents. We argue that policy documents primarily function as a category of boundary objects that Star and Griesemer (1989) call the “ideal type.” In other words, platform policies outline an ideal vision of the platform, userbase, and even the broader world transformed through platformization. This normative vision creates a script that users encounter and platforms promote.

Different policies outline different ideals, as platform policies include a range of formal and informal documents. Some are legally binding while others function as guidelines (Pater et al., 2016). In this article, we focus on three policies that fulfill different roles. First, the TOS outline the ideal relationship between the user and the platform. The TOS is a legal agreement between the user and service provider that lays out the rights and responsibilities of both parties. Prior research has discussed the inclusion and exclusion of specific values in TOS documents, especially freedom of expression, a value that is deeply rooted in the American Constitution (Klonick, 2017). Similarly, DeNardis and Hackl’s (2015) analysis of corporate policies focuses on the areas of privacy, expression, and interoperability to show how platform policies use these central values as a justification for self-governance.

Privacy Policies outline the ideal treatment of personal data. Such policies have been widely researched, with many scholars highlighting the challenge of studying privacy given its contested interpretations (Epstein et al., 2014). Social media users have a complex relationship with the concept of privacy, where “people appear to want and value

privacy, yet simultaneously appear not to value or want it” (Nissenbaum, 2010: 104). Despite the contradictions, privacy has been often depicted as a human right and acknowledged as a value (Epstein et al., 2014). Researchers have examined how users understand and interact with Privacy Policies (Quinn et al., 2019), generally finding that users understand and define privacy differently than institutions.

Finally, Community Guidelines outline the ideal experience of platforms as a community, showing how people should express themselves and interact with others. Community Guidelines are the most approachable or “user-friendly” documents (Maddox and Malson, 2020). As the name of the policy suggests, the concept of “community” is pivotal in social media-related discourse (Baym, 2000). Like TOS documents, Community Guidelines emphasize the value of free expression (Gillespie, 2018) and epitomize “Silicon Valley ideals” (Maddox and Malson, 2020).

While these studies are informative in demonstrating the strong value-orientation of social media policy documents, their focus is mainly on freedom of expression and privacy. Aiming to expand the range of values investigated outlook, we set out to explore the following questions:

RQ1. What values are invoked in the TOS, Privacy Policies, and Community Guidelines of social media platforms?

RQ2. What actors are these values associated with?

RQ3. Do different platforms convey different values in their policy documents?

Methods

We combined several approaches to provide a comprehensive account of the values foregrounded in platform policies. First, we charted a bird’s eye view of value terms through a quantitative analysis of the words composing these documents. As detailed below, this analysis was instructive in identifying the main realms that platforms emphasize, yet limited in its ability to identify deeper discursive mechanisms and overarching principles. To this end, we conducted an independent qualitative analysis of the policies that strove to identify how platforms assign responsibility to various stakeholders and which values are used as governing principles to construct the ideal platform and user.

Retrieval of platform policies

Our corpus includes the Privacy Policy, TOS, and Community Guidelines of Facebook, Instagram, YouTube, TikTok, and Twitter. The main consideration guiding the selection of the first four platforms was their international popularity (Statista, 2022). While instant messaging apps such as WhatsApp are also widely used, we excluded them from this analysis since they involve different governance considerations given their role as intermediators between primarily known contacts (Gillespie, 2018). In addition, we decided to include Twitter due to its long-lasting political significance across many contexts. While all five share a “techno-commercial architecture [. . .] rooted in neoliberal market value” (Van Dijck, 2020), they differ in their ownership structure. Alphabet, the

Table 1. Overview of the platform policy documents by words.

Platform/Document	Facebook	Instagram	TikTok	Twitter	YouTube
Privacy Policy	4235 words	4215 words	3040 words	5501 words	4134 words
Terms of Service	4143 words	2450 words	7499 words	3038 words	4186 words
Community Guidelines	14,283 words	1402 words	3305 words	11,560 words	9729 words

owner of YouTube, has a different corporate setup with more diverse businesses compared with Meta, ByteDance, and Twitter. The documents were retrieved in June 2020 and refer to US jurisdictions (see discussion of this limitation below). As the Community Guidelines of Facebook, Twitter, YouTube, and Instagram consist of shorter versions to enhance readability, we followed “learn more” hyperlinks when presented in these policies and combined all texts into a single document. The resulting dataset consists of 15 documents, ranging from 1402 to 14,283 words (see Table 1).

Identifying and clustering value terms

To assess which values are explicitly mentioned in platform policies, we used WordStat to create a lemmatized list of all the words in the corpus sorted by frequency. After lemmatization, several words were combined manually given their shared root. In the next step, we created a smaller corpus of words that appear at least 10 times in the documents ($n=820$). We excluded “community” and “privacy” from this corpus since they are included in the names of the policies and often invoked to refer to them. Applying the conceptualization of values as principles by Heinich (2020), we counted a word as a value if it (1) could guide decisions or behavior, (2) has positive associations, and (3) could be applied to multiple contexts. We included words that could *potentially* guide behavior (e.g. “loyalty”), yet did not determine in advance whether the words were indeed used as guiding principles in the documents. We followed the practices of consensual qualitative research (see Hill et al., 1997; Spangler et al., 2012) to evaluate the applicability of the terms to our definition of values. This protocol allows for systematic evaluation of complex data, with researchers separately applying codes and reaching agreement through deliberation. This phase yielded a list of 98 values that we trimmed to 66 after combining verbs and nouns (e.g. “choose” and “choice”).

To make sense of the list of values, the three researchers reevaluated it to detect clusters, or groups of similar values. Following modified principles of the grounded theory approach (see Kelle, 2007), which combines an inductive examination of data with existing theoretical literature, we evaluated the words comparatively against each other in light of existing value schemes (Christen et al., 2016; Schwartz, 2012). This process led us to organize the value terms into 10 clusters: engagement, safety, power, agency, expression, improvement, effort, care, awareness, and responsibility.

Qualitative content analysis

Alongside our word-based analysis, we wanted to uncover deeper patterns underpinning the construction of values in these documents. To this end, we conducted a thematic

analysis focusing on the different types of actors that the policies address: the platform, general users, and business users. For each actor, we coded for *desirable* and *undesirable* behavior. Additional codes concerned phrases that address all actors (e.g. “we,” “society”) and external actors (e.g. the law, researchers). Here again, we followed the practices of consensual qualitative research, using MaxQDA to aid us in this process. Once we finished annotating the texts according to our scheme, we reevaluated the results according to Heinrich’s conceptualization of values as principles, aiming to identify broad standards that govern the depiction of the “ideal” user and platform across policies. This led us to identify five core values that serve as governing principles in platform policies.

To explore the implications of our findings in relation to the aforementioned literature on stakeholders and commercial interests, we reanalyzed the whole corpus focusing on the five core principles we identified. At this phase, we explored possible tensions arising from the ways policies assign responsibility for enacting values to different stakeholders.

Results

Overall, our analysis revealed a consistent set of value terms and organizing logics across platform policies. In what follows, we first survey the 10 prominent clusters of values that emerged from the word-based analysis, paying particular attention not only to the values present but also to some “public values” that did not feature prominently in the policies. Second, we present the outcome of our qualitative analysis, describing five core values that serve as governing principles for the platforms.

Clusters of platform values

The 10 value clusters we identified in the policy documents are engagement, safety, power, agency, expression, improvement, effort, care, awareness, and responsibility (see Table 2). *Engagement* is the largest cluster and includes seven value terms associated with participation, sharing, or connection. This is followed by the cluster we call *improvement*, which brings together values such as impact, change, and development. The third cluster is *safety*, gathering seven terms associated with security. The *care* cluster bundles six values related to help or support, terms that are relatively common in the documents. This is followed by *expression*, clustering seven values connected to communication and authenticity. *Awareness* contains eight values associated with knowledge and information. *Power* incorporates eight values that are mainly artifacts of legal frameworks. *Agency* clusters four values related to freedom and choice, while *effort* covers four values related to terms like commitment. *Responsibility*, which only includes two words with a shared root, is the smallest cluster in the corpus.

The set of value terms is consistent across platforms. All 10 clusters appear in the policies of the five platforms, yet there are some minor differences in invoking specific terms. Most values (40 out of 66) appear in all five policies, 18 are shared by four policies, 5 appear in three policies, and only 3 appear in two or fewer platform policies. For instance, *advocate* is mainly mentioned by Facebook, *civic/civil* by Twitter, and *trust* by TikTok.

Table 2. The 10 value clusters and frequencies of associated value terms.

Word	Total	Facebook	Instagram	TikTok	Twitter	YouTube
Engagement						
Share	382 (0.46%)	130 (0.57%)	66 (0.82%)	28 (0.20%)	120 (0.60%)	38 (0.21%)
Engage*	140 (0.17%)	76 (0.34%)	7 (0.09%)	5 (0.04%)	44 (0.22%)	8 (0.04%)
Connect*	65 (0.08%)	29 (0.13%)	15 (0.19%)	10 (0.07%)	8 (0.04%)	3 (0.02%)
Relate	67 (0.08%)	13 (0.06%)	8 (0.10%)	15 (0.11%)	22 (0.11%)	9 (0.05%)
Unite	31 (0.04%)	6 (0.03%)	3 (0.04%)	6 (0.04%)	16 (0.08%)	0 (0.00%)
Participate*	48 (0.06%)	14 (0.06%)	0 (0.00%)	6 (0.04%)	18 (0.09%)	10 (0.06%)
Civic*	36 (0.04%)	6 (0.03%)	0 (0.00%)	2 (0.01%)	28 (0.14%)	0 (0.00%)
Improvement						
Change	88 (0.11%)	21 (0.09%)	11 (0.14%)	8 (0.06%)	14 (0.07%)	34 (0.19%)
Promote	111 (0.13%)	29 (0.13%)	4 (0.05%)	24 (0.17%)	33 (0.16%)	21 (0.12%)
Encourage	77 (0.09%)	16 (0.07%)	5 (0.06%)	13 (0.09%)	20 (0.10%)	23 (0.13%)
Develop*	58 (0.07%)	32 (0.14%)	5 (0.06%)	3 (0.02%)	12 (0.06%)	6 (0.03%)
Improve	33 (0.04%)	13 (0.06%)	5 (0.06%)	3 (0.02%)	4 (0.02%)	8 (0.04%)
Safety						
Protect*	145 (0.18%)	35 (0.15%)	10 (0.12%)	14 (0.10%)	38 (0.19%)	48 (0.27%)
Safe*	127 (0.15%)	35 (0.15%)	14 (0.17%)	16 (0.12%)	25 (0.12%)	37 (0.20%)
Secure*	73 (0.09%)	10 (0.04%)	9 (0.11%)	11 (0.08%)	18 (0.09%)	25 (0.14%)
Care						
Help	166 (0.20%)	48 (0.21%)	30 (0.37%)	5 (0.04%)	38 (0.19%)	45 (0.25%)
Give	62 (0.07%)	24 (0.11%)	9 (0.11%)	4 (0.03%)	7 (0.03%)	18 (0.10%)
Believe	68 (0.08%)	14 (0.06%)	6 (0.07%)	3 (0.02%)	31 (0.15%)	14 (0.08%)
Support	58 (0.07%)	25 (0.11%)	8 (0.10%)	11 (0.08%)	10 (0.05%)	4 (0.02%)
Expression						
Create*	152 (0.18%)	45 (0.20%)	23 (0.29%)	14 (0.10%)	24 (0.12%)	46 (0.25%)
Communicate*	61 (0.07%)	19 (0.08%)	13 (0.16%)	9 (0.07%)	16 (0.08%)	4 (0.02%)
Express*	72 (0.09%)	30 (0.13%)	5 (0.06%)	7 (0.05%)	27 (0.13%)	3 (0.02%)
Speak*	54 (0.07%)	34 (0.15%)	4 (0.05%)	4 (0.03%)	3 (0.01%)	9 (0.05%)
Awareness						
Learn	113 (0.14%)	38 (0.17%)	28 (0.35%)	1 (0.01%)	30 (0.15%)	16 (0.09%)
Understand	74 (0.09%)	23 (0.10%)	9 (0.11%)	9 (0.07%)	12 (0.06%)	21 (0.12%)
Know	47 (0.06%)	19 (0.08%)	7 (0.09%)	7 (0.05%)	8 (0.04%)	6 (0.03%)
Aware*	38 (0.05%)	18 (0.08%)	3 (0.04%)	11 (0.08%)	1 (0.00%)	5 (0.03%)
Think	31 (0.04%)	8 (0.04%)	4 (0.05%)	0 (0.00%)	3 (0.01%)	16 (0.09%)
Power						
Control	76 (0.09%)	13 (0.06%)	9 (0.11%)	11 (0.08%)	23 (0.11%)	20 (0.11%)
Comply	41 (0.05%)	8 (0.04%)	0 (0.00%)	13 (0.09%)	8 (0.04%)	12 (0.07%)
Order	46 (0.06%)	12 (0.05%)	3 (0.04%)	6 (0.04%)	15 (0.07%)	10 (0.06%)
Authorize	36 (0.04%)	6 (0.03%)	0 (0.00%)	10 (0.07%)	14 (0.07%)	6 (0.03%)
Govern	23 (0.03%)	5 (0.02%)	3 (0.04%)	5 (0.04%)	7 (0.03%)	3 (0.02%)
Lead	31 (0.04%)	8 (0.04%)	0 (0.00%)	7 (0.05%)	9 (0.04%)	7 (0.04%)
Agency						
Choose*	76 (0.09%)	20 (0.09%)	16 (0.20%)	19 (0.14%)	12 (0.06%)	9 (0.05%)
Effort						
Commit*	57 (0.07%)	30 (0.13%)	5 (0.06%)	12 (0.09%)	1 (0.00%)	9 (0.05%)
Responsibility						
Responsible*	59 (0.07%)	9 (0.04%)	5 (0.06%)	18 (0.13%)	9 (0.04%)	18 (0.10%)

*Words from the same root were combined in these entries (e.g. choice and choose).

Note: Only values that appeared more than 30 times were included in the table. Percentages in parentheses refer to the number of occurrences of value terms divided by the total number of words in the relevant documents.

Interestingly, our quantitative results show that “public values” invoked in the literature associated with platform rhetoric such as democracy (Vaidhyathan, 2018), transparency (Van Dijck et al., 2018), and fairness (Cath et al., 2020) are not just “increasingly compromised” (Van Dijck, 2020) but largely absent from the policy documents. None of these terms or their synonyms made the threshold of 10 appearances in the documents. *Transparency* was invoked only eight times (all related to the publication of “transparency reports”), *freedom* came up four times, and *democracy*, *accountability*, *neutrality*, and *equality* did not appear at all.

Core values

The qualitative content analysis revealed that platforms share several organizing logics, or underlying normative justifications given for policy decisions. We identified five core values associated with these justifications: *expression*, *community*, *safety*, *choice*, and *improvement*. As we elaborate below, with only one exception, these values were invoked across all five platforms, with some variation in their framing. While Facebook and Instagram, which are owned by the same corporation, were closer to each other in the framing of some of these values, they did not differ dramatically from the other three platforms. Moreover, the division of responsibility for the execution of values was similar across platforms, with comparable roles assigned to users and platforms.

Expression. When conceptualized as a value, expression means that people should articulate their thoughts, opinions, and emotions. The notion of *free* expression, common in the policies, suggests that people should be also able to voice their opinions with minimal restrictions. This is closely connected to the principles of free speech and the metaphorical “marketplace of ideas” (Maddox and Malson, 2020), which is often used on social media platforms to signal the pursuit of “democratic self-governance” (Jones, 2018) and thus provides a justificatory framework for content moderation decisions (Klonick, 2017).

Our analysis of policy documents strongly demonstrates this duality. On one hand, all platforms endorse the value of expression, valorizing the creation of environments in which users can express themselves freely and openly (Facebook, Twitter), authentically (Instagram), or creatively (TikTok, YouTube). This is epitomized through statements such as “Free expression is a human right—we believe that everyone has a voice, and the right to use it” (Twitter, 2021). On the other hand, expression is often invoked in policies as a justification for limiting speech. Facebook (2021a), for example, states: “Our commitment to expression is paramount, but we recognize the internet creates new and increased opportunities for abuse. For these reasons, when we limit expression, we do it in service of one or more [sic] the following values [Authenticity, Safety, Privacy, Dignity].” Thus, all platforms endorse the value of expression while presenting restrictions as necessary to create an environment where users feel comfortable and safe to express themselves.

The responsibility for enacting the value of expression is shared between users and platforms. Through constant requests to *engage*, *share*, and *communicate*, the ideal user is framed as inherently expressive: someone who regularly shares content and interacts

with others on social media. Yet users are also asked to restrict their expression, as the policies explain that some people will only feel comfortable sharing content in the right kind of environment, characterized by positive and “authentic” (Facebook, 2021a) content. While users should do their part to consider others and make good decisions regarding self-expression, platforms are responsible for creating a safe and conducive environment. Policy documents, especially Community Guidelines, describe the role of platforms in enabling good expression through the restriction of bad content and potentially harmful behavior, as well as the development of tools that detect and moderate negative content such as bullying or hate speech.

Community. Given its longstanding importance for thinking about the Web (Baym, 2000), it is unsurprising that community appears as a core justification for social media platforms. Rheingold (2000) defines virtual communities as emerging when “enough people carry on those public discussions long enough, with sufficient human feeling, to form webs of personal relationships in cyberspace” (p. 5). Community can therefore be defined in this context as a set of intimate relations developed or sustained through communication technologies. As such, it acts as a value both in terms of something that people consistently evaluate positively and as a guiding principle for the evaluation of platforms and users. All platforms except Twitter (see below) frame the formation and protection of community as a primary purpose of social media.

While some forms of expression are framed negatively, community is univocally framed as good, with platforms using phrases such as “fostering a safe and supportive community for everyone” (Instagram, 2021a), or “a global community where people can create and share, discover the world around them, and connect with others across the globe” (TikTok, 2021a). Policy documents do not specifically define community, despite their frequent use of the term. In line with our definition above, they tend to invoke the general idea of maintaining ties with other people and the platform itself. However, they introduce fuzziness to the idea of community by unbounding it to include an unlimited set of potential bonds between all users of a given platform around the globe. In this expansive sense, community is not defined as a specific group of people that share the same interests but rather by the mere use of a platform.

The responsibility for bringing community into existence lies particularly on users, as they are the ones expected to actively “foster” community. Users should strive to share ideas, opinions, and content with others—basically, express themselves—in order to create and maintain a community. The policy documents note that platforms can facilitate community by connecting users through their technical infrastructures, recommending relevant people and content. Beyond recommendations, platforms also position their prohibitions on behavior as in service of community, when asking, for example, to “Respect everyone on Instagram, don’t spam people or post nudity” (Instagram, 2021a).

The platforms share similar ideas about the importance of building community, being an active member, and feeling close to people that share the same interests. While Twitter’s “rules” are similar to Community Guidelines, Twitter does not invoke the term “community” and instead describes its purpose as “serv[ing] the public conversation.” However, Twitter policies do reflect the goal of creating spaces in which people are free

to exchange, communicate, and express themselves, similar to how community is framed by other platforms.

Safety. As a core value, safety can be broadly defined as freedom from danger or risk. In the context of social media, users perceive safety as closely entwined with security, privacy, and community, highlighting it as pivotal to their daily experiences (Redmiles et al., 2019). Various policy documents stress the importance of safety, framing it as a prerequisite for expression and community. According to TikTok (2021a), “feeling safe is essential to helping people feel comfortable with expressing themselves openly and creatively.” As such, safety serves as a justification for setting boundaries on permissible content, users, and behavior.

The responsibility for enacting and maintaining safety is shared between users and platforms, but each actor has different tasks. Users are supposed to take advantage of the tools created by platforms and educate themselves about what they can do to “protect their privacy.” Privacy and safety are intertwined in the policies, with safety serving as a higher-level value, while privacy is closely tied to individual choices that ultimately lead one to feel safer. For platforms, safety functions as a justification for moderating harassment, hate speech, and disinformation. Examples of threats to safety include “promoting engagement in child sexual exploitation” (Facebook, 2021b) or content that is “gratuitously shocking, sadistic, or excessively graphic” (TikTok, 2021b). This responsibility is often articulated alongside the development and deployment of automated technological solutions for content moderation.

Platforms share the value of safety but differ in how they present it. Facebook, YouTube, and TikTok stress details about what users are not allowed to do and what kind of content is generally allowed on the platform. All three platforms provide specific definitions of unacceptable behavior that can lead to physical harm. The use of automated content moderation systems to create a safer platform and serve the greater good is prevalent in the policies of Instagram, Twitter, and YouTube. Overall, the meaning of safety is left ambiguous and open to interpretation even as it is consistently framed as a vital part of platforms.

Choice. Widely invoked in platform policies, choice signifies that individuals should be free to pick options that align with their interests. Through interface design, account personalization, and informational resources, platforms provide options and frame the ideal user as someone who makes active and informed choices that align with their preferences. The appeal to personal autonomy, or the “freedom to have and make choices,” has long been cherished by liberal democracies and celebrated as a defining feature of the Web (Graham and Henman, 2019). Following this ethos, policy documents repeatedly promote choices around personal information as a source of empowerment granted by the platform to users. Twitter connects the idea of choice to control: “You’re in control. Even as Twitter evolves, you can always change your privacy settings. The power is yours to choose what you share in the world” (Twitter, 2021).

Although the user has the “power to choose,” the platform plays a major role in structuring the process, formatting options, establishing default settings, and nudging users toward particular paths through design. Despite the rhetorical promise, this has the effect

of significantly narrowing the experience of choice, a dynamic common on the commercial Web. As Graham and Henman (2019) report, “Commercial websites delimit the scale of choice on offer to different degrees whilst at the same time producing the feeling of ‘global’ and ‘informed’ choice” (p. 2020). In the context of social media, users navigate a range of carefully delimited choices, from the initial choice to accept the TOS to join the network, to various choices around personal account settings, to everyday choices about how to interact with others on the platform. Of these, policy documents focus on personal settings for sharing data with the platform and/or with the public. In order to move away from the default of sharing, users have to actively seek out these options (as the platforms suggest, to “learn more”).

Platforms facilitate choice by presenting options for users and incorporating potential choices in their design, but the responsibility for realizing choice ultimately falls onto the user. In policies of all five platforms, choice is connected to decisions about sharing content, personal information, or association with third parties. Ultimately, even if individuals do not change any of their default settings, according to the policies, accepting the default setting is still a choice and users are the ones in control. The underlying implication is that individuals are responsible for the consequences of their choices.

Improvement. When articulated as a value, improvement often stands in as a synonym for growth and the creation of economic value (Arvidsson and Colleoni, 2012). This framing fits within the broader trajectory of technological progress, an underlying principle of Western societies (Nye, 2003). Social media companies have adopted this imperative of growth through improvement, with platforms constantly striving to “scale up” in terms of available features and active userbase, and “scale out” to become central actors of private and public life (Hallinan, 2021).

The goal of growth often involves reaching out to other businesses rather than users. Examining the networked partnerships of popular social media platforms, Van der Vlist and Helmond (2021) show that these economic relationships are essentially a “global partner ecosystem” that builds an extensive organizational infrastructure and creates “nodes of power” (p. 10) crucial for platforms’ growth. Such interdependency pertains also to relationships between corporations. Caplan and boyd (2018) demonstrate how data-driven practices of big tech corporations such as Facebook have become institutionalized and their “algorithmic logics” function as administrative mechanisms that organize relationships, resulting in a homogeneous media landscape. In this interdependent, institutionalized ecosystem, “improvements” achieved by one platform are likely to be quickly adopted by other platforms—a rising tide that raises all boats.

In the policy documents, the value of improvement is emphasized in relation to the technological infrastructures of platforms, from the design of recommendation systems to the integration of third parties. Platforms promise to create a “better,” more “seamless” (Facebook, 2021a) user experience and to show content closer to an individual’s interests, or “content that matters” (Facebook, 2021a). Using this constant positive framing of their actions, platforms emphasize their role in bettering people’s lives.

The responsibility for enacting improvement almost exclusively falls to the platforms. Platforms use this value to highlight long-term goals like making the platform “safer” (Instagram, 2021b) or offering a “better” service. Despite the importance of improvement,

the policies do not suggest that this value is something most users need to act on. However, advertisers can participate in improvement by virtue of receiving ever more granular audience data. While the responsibility for improvement lies predominantly on the side of platforms, the outcome is supposed to serve everyone. With this value, platforms position themselves as enablers of societal change.

Discussion

We examined the discourse of values in social media platform policies from two directions, combining a quantitative analysis of value-related terms with a qualitative investigation of the principles platforms draw upon to justify their policies. At first glance, the two approaches seem to overlap only partially in their accounts of platform values. However, a closer look reveals a complementary relationship between the two methods, illustrating both the complexity of the value concept and the ways in which platforms use values as discursive resources.

Some of the categories identified through the analysis of value-related terms map clearly onto the governing principles we identified. These include *improvement*, *safety*, *expression*, and *choice*. An examination of the four other value-related clusters that we did not identify as core principles—*effort*, *care*, *responsibility*, and *awareness*—shows that terms associated with them are often invoked as a means to achieve the core values. Thus, for instance, responsible behavior contributes to a sense of safety and awareness is a prerequisite for making informed choices. The group of terms we bundled as relating to *power* includes expressions such as “comply” and “authorize” which are anchored in legal discourse and reflect, more than the other clusters, the aforementioned use of policies as “boundary objects” that speak to divergent audiences (Arun, 2022). Finally, while we did not identify *engagement* as a core value, many of the ideas reflected in what we frame as the *community* and *expression* values relate to user engagement, both in the sense of connecting to others and of being active on the platform.

Although platforms vary in the comprehensiveness of their policies (Jiang et al., 2020) and the specific behaviors they discuss (Pater et al., 2016), our analysis reveals that they employ a surprisingly consistent set of justifications that are broad and, at least potentially, compatible with the public good (Moore, 1995). As governing principles, there is nothing that necessarily ties the core values of *improvement*, *safety*, *expression*, *community*, and *choice* to the context of social media. Indeed, a part of the strength of these values is that they can apply to a range of contexts, from interpersonal to corporate, public to private. In so doing, these principles seem to suggest a broader discourse of values than the cyberlibertarian-inflected “Silicon Valley ideals” previously identified in Community Guidelines policies (Maddox and Malson, 2020), especially the invocation of *community* and *safety*. However, despite the fact that platforms promote values that have the potential to serve the public good, they are placed within a commercial framework that allows these corporations to strategically reinterpret values to suit their own interests. In other words, platform policies tie both public and private interests to the same principles.

This applies to each governing principle. *Expression* is presented as both the reason for a platform’s existence and an imperative for users, whether engaging in political

speech framed as a human right or in creative acts of communication. At the same time, every piece of expression is also a part of the business model and a mode of revenue generation. Thus, for social media platforms, expression is simultaneously social and profitable. In an economic environment where the *active* participation of users is key to their corporate assetization (Birch et al., 2021), the more participation in the marketplace, the better.

A similar situation arises with the principle of *community*. Whereas platforms stress the public benefit of creating a community where people feel safe to connect and express themselves, invoking one of the longstanding appeals of digital technologies (Baym, 2000), such connections implicitly establish a relationship with the platform itself. Simply put, platforms want users to stay on the platform, and every relationship and positive experience creates network dependencies (Griggio et al., 2022), making the bond between user and platform harder to cut.

Safety provides the justification for a number of active measures, from data protection policies to content moderation to choosing privacy settings. Such initiatives are framed as serving users and the broader societal good, even as the notion of harm motivating safety initiatives is strategically vague and emphasizes individual misconduct rather than structural forms of violence (DeCook et al., 2022). Furthermore, such measures also serve companies' interests by ensuring legal compliance and helping them avoid government regulations and external interventions.

In the repeated promotion of *choice*, platforms emphasize the autonomy or freedom of users. The appeals to "learn more" and make choices construct the ideal user as an informed, active subject, a longstanding move in platform policy documents (Stein, 2013) and an inheritance of the "notice and choice" legal framework for privacy protection established in the United States in the 1970s (Obar and Oeldorf-Hirsch, 2020). Despite the rhetorical celebration of choice, the version of user autonomy outlined in the policy documents is exceedingly limited. In contrast, the autonomy of platforms themselves, exercised through the creation of default interfaces and set menus of choices offered to users, remains unaddressed.

Finally, the value of *improvement* fits within a larger narrative of progress. The documents often connect improvement to technological development, with a plethora of assertions about providing "better" services to all users. Platforms use positive terms such as "bigger" and "more personalized" to sugarcoat economic initiatives to both save money (e.g. through automated systems) and make money (e.g. through personalized ads). However, where the public statements of social media executives like Mark Zuckerberg take the next step of connecting technological improvement with social transformation and making a "better world" (Haupt, 2021), policies promote a narrower version of improvement concerned with user-experience and shareholder value.

In bundling together public and economic interests into the same principles, platform policy documents reproduce a broader discursive strategy of using positively charged value terms to promote commercial interests (Epstein et al., 2014; Hallinan et al., 2022; John, 2016; Van Dijck, 2013). Yet beyond this bundling of interests, platforms further delimit or dilute the potential public impact of value commitments by strategically assigning and largely offloading responsibility for enactment of the said values. Among

the five governing principles, the only one that platforms take full responsibility for is *improvement*. The enactment of the remaining values requires the interaction and coordination of both users and platforms, with platforms providing tools that users are expected to utilize in specific ways. As such, the ultimate responsibility for realizing the values of *expression*, *community*, *safety*, and *choice* lies with users, articulating a vision of platform governance that significantly depends on self-governance (DeNardis and Hackl, 2015). Thus, even though the governing principles of platform policies seem to extend beyond the cyberlibertarian vision (Maddox and Malson, 2020), with professed commitments to *safety* and *community*, the assignment of responsibility is largely individualistic.

The limited public value of the governing principles is reinforced through a consideration of the values that are absent or rarely invoked in these documents, including *democracy*, *accountability*, *transparency*, and *neutrality*. While such values are prominent in the academic literature and the discourse of policymakers, they are notably not a part of policy justifications. Even *transparency*, the only word invoked with any frequency among the group, appeared as an object rather than a principle, a type of report created and published by companies to appease policymakers and avoid further regulation. In contrast to transparency, which can be “downgraded” to technical measures without much difficulty, invoking values such as *democracy* or *accountability* may pose a risk for commercial interests, as they invoke a broad set of considerations that may open companies to unwanted scrutiny or regulatory interference.

Conclusion

Our mixed-methods analysis of social media platform policies offers observations about the values invoked in these documents and their organizing logics. Yet our study has two main limitations that will hopefully be addressed in future studies. First, we only focused on US-based policies. To create a holistic account of the values constructing platform policies, researchers should investigate non-US policies and consider their respective contexts. Second, while this study focuses on the exploration of values in policy documents, it does not examine if or how companies promote these values in other venues. To further understand the construction of platform values, future research could investigate whether and how values that are prevalent in policy documents manifest in public statements, moderation practices, and platform design.

Notwithstanding these limitations, our study sheds light on how platforms perform boundary work through the invocation of values in their policies. While they shy away from some public values that may not align with commercial interests, platforms highlight values that allow them to have their cake and eat it too. Invoking *improvement*, *choice*, *community*, *expression*, and *safety* conveys platforms’ commitment to a greater good while simultaneously protecting their revenue streams. Moreover, by assigning much of the responsibility for enacting such values onto users, platforms downplay their role in promoting them. In this sense, the core values in these policies reflect systematic and likely enduring tensions between the economic and moral meanings of value(s) for social media platforms.

Acknowledgements

We are indebted to the anonymous reviewers for their collegial and insightful comments on this manuscript. We also thank CJ Reynolds, Tommaso Trillò, Bumsoo Kim, Maximilian Overbeck, Nicholas John, Dmitry Epstein, and Anat Ben-David for their valuable feedback and assistance.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant Agreement No. 819004).

ORCID iDs

Rebecca Scharlach  <https://orcid.org/0000-0003-3033-5795>

Blake Hallinan  <https://orcid.org/0000-0002-4696-8290>

Limor Shifman  <https://orcid.org/0000-0002-2616-9491>

References

- Ahn S, Baik JS and Krause CS (2022) Splintering and centralizing platform governance: how Facebook adapted its content moderation practices to the political and legal contexts in the United States, Germany, and South Korea. *Information, Communication & Society*. Epub ahead of print 29 August. DOI: 10.1080/1369118X.2022.2113817.
- Andersson Schwarz J (2022) Knowledge-making on techno-commercial platforms: the example of Facebook. In: Vossen P and Fokkens A (eds) *Creating a More Transparent Internet: The Perspective Web*. Malden, MA: Cambridge, pp. 129–142.
- Arun C (2022) Facebook's faces. *Harvard Law Review*. Available at: <https://harvardlawreview.org/2022/03/facebooks-faces/> (accessed 7 December 2022).
- Arvidsson A and Colleoni E (2012) Value in informational capitalism and on the internet. *The Information Society* 28(3): 135–150.
- Barrett B and Kreiss D (2019) Platform transience: changes in Facebook's policies, procedures, and affordances in global electoral politics. *Internet Policy Review* 8(4): 1446.
- Baruh L and Popescu M (2017) Big data analytics and the limits of privacy self-management. *New Media & Society* 19(4): 579–596.
- Baym NK (2000) *Tune In, Log on: Soaps, Fandom, and Online Community*. Thousand Oaks, CA: SAGE.
- Birch K, Cochrane D and Ward C (2021) Data as asset? The measurement, governance, and valuation of digital personal data by Big Tech. *Big Data & Society* 8(1): 1–15.
- Boltanski L and Thévenot L (2006) *On Justification: Economies of Worth* (Princeton Studies in Cultural Sociology). Princeton, NJ: Princeton University Press.
- Caplan R and boyd d (2018) Isomorphism through algorithms: institutional dependencies in the case of Facebook. *Big Data & Society* 5(1): 1–12.
- Cath C, Latonero M, Marda V, et al. (2020) Leap of FATE: human rights as a complementary framework for AI policy and practice. In: *Proceedings of the 2020 conference on fairness, accountability, and transparency*, New York, 27 January. New York: ACM Press.
- Christen M, Narvaez D, Tanner C, et al. (2016) Using thesauruses as a heuristic for mapping values. *Cognitive Systems Research* 40: 59–74.
- DeCook JR, Cotter K, Kanthawala S, et al. (2022) Safe from “harm”: the governance of violence by platforms. *Policy & Internet* 14(1): 63–78.

- DeNardis L and Hackl AM (2015) Internet governance by social media platforms. *Telecommunications Policy* 39(9): 761–770.
- Epstein D, Roth MC and Baumer EPS (2014) It's the definition, stupid! Framing of online privacy in the internet governance forum debates. *Journal of Information Policy* 4: 144.
- Facebook (2015) Facebook's 5 Core Values. Available at: https://www.facebook.com/facebook-careers/photos/?tab=album&album_id=1655178611435493 (accessed 14 January 2022).
- Facebook (2021a) Community standards. Available at: <https://www.facebook.com/community-standards> (accessed 31 October 2021).
- Facebook (2021b) Terms of service. Available at: <https://www.facebook.com/legal/terms> (accessed 14 January 2022).
- Fiesler C and Hallinan B (2018) 'We are the product': public reactions to online data sharing and privacy controversies in the media. In: *Proceedings of the 2018 CHI conference on human factors in computing systems—(CHI'18)*, Montreal, QC, Canada, pp. 1–13. ACM Press.
- Gandy OH (2011) The political economy of personal information. In: Weingast BR and Wittman DA (eds) *The Handbook of Political Economy of Communications*. London: John Wiley & Sons, pp. 436–457.
- Gillespie T (2017) Regulation of and by platforms. In: Poell T, Marwick A and Burgess J (Eds) *The SAGE Handbook of Social Media*. Los Angeles, CA: SAGE, pp. 417–456.
- Gillespie T (2018) *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. New Haven, CT: Yale University Press.
- Gillespie T, Aufderheide P, Carmi E, et al. (2020) Expanding the debate about content moderation: scholarly research agendas for the coming policy debates. *Internet Policy Review* 9(4): 1512.
- Gorwa R (2019) What is platform governance? *Information, Communication & Society* 22(6): 854–871.
- Graham T and Henman P (2019) Affording choice: how website designs create and constrain "choice." *Information, Communication & Society* 22(13): 2007–2023.
- Griggio CF, Nouwens M and Klokmose CN (2022) Caught in the network: the impact of WhatsApp's 2021 privacy policy update on users' messaging app ecosystems. In: *Proceedings of the 2022 CHI conference on human factors in computing systems (CHI'22)*, New York, 28 April, pp. 1–23. New York: ACM Press.
- Hallinan B (2021) Civilizing infrastructure. *Cultural Studies* 35: 707–727.
- Hallinan B, Scharlach R and Shifman L (2022) Beyond neutrality: conceptualizing platform values. *Communication Theory* 32(2): 201–222.
- Haupt J (2021) Facebook futures: Mark Zuckerberg's discursive construction of a better world. *New Media & Society* 23(2): 237–257.
- Heinich N (2020) Ten proposals on values. *Cultural Sociology* 14(3): 213–232.
- Hill CE, Thompson BJ and Williams EN (1997) A guide to conducting consensual qualitative research. *The Counseling Psychologist* 25(4): 517–572.
- Instagram (2021a) Community guidelines. Available at: <https://www.facebook.com/help/instagram/477434105621119/> (accessed 13 April 2021).
- Instagram (2021b) Data policy. Instagram Help Center. Available at: <https://help.instagram.com/519522125107875> (accessed 31 October 2021).
- Jiang J, Middler S, Brubaker JR, et al. (2020) Characterizing community guidelines on social media platforms. In: *Conference companion publication of the 2020 on computer supported cooperative work and social computing*, 17 October, pp. 287–291. ACM Press.
- Joas H (1997) *Die Entstehung der Werte* [The Genesis of Values]. 1st ed. Frankfurt am Main: Suhrkamp.
- John NA (2016) *The Age of Sharing*. Cambridge, MA: Polity.

- Jones RL (2018) Can you have too much of a good thing? The modern marketplace of ideas. *Missouri Law Review* 83: 19.
- Kelle U (2007) The development of categories: different approaches in grounded theory. In: Charmaz K (ed.) *The SAGE handbook of grounded theory*. London: SAGE, pp. 191–213.
- Klonick K (2017) The new governors: the people, rules, and processes governing online speech. *Harvard Law Review* 131(6): 1598–1670.
- Kopf S (2022) Corporate censorship online: vagueness and discursive imprecision in YouTube’s advertiser-friendly content guidelines. *New Media & Society*. Epub ahead of print 8 February. DOI: 10.1177/14614448221077354.
- Lamont M (2012) Toward a comparative sociology of valuation and evaluation. *Annual Review of Sociology* 38(1): 201–221.
- Leurs K and Zimmer M (2017) Platform values: an introduction to the #AoIR16 special issue. *Information, Communication & Society* 20(6): 803–808.
- Maddox J and Malson J (2020) Guidelines without lines, communities without orders: the marketplace of ideas and digital manifest destiny in social media platform policies. *Social Media + Society*. Epub ahead of print 19 June. DOI: 10.1177/2056305120926622.
- Martin JL and Lembo A (2020) On the other side of values. *American Journal of Sociology* 126(1): 52–98.
- Moore MH (1995) *Creating Public Value: Strategic Management in Government*. Cambridge, MA: Harvard University Press.
- Mosseri A (2021) Tweet: we’re not neutral. @Mosseri. Available at: <https://twitter.com/mosseri/status/1348713108127309824> (accessed 12 July 2021).
- Nakamura L (2021) Women of color and the digital labor of repair. *Microsoft Research*. Available at: <https://www.microsoft.com/en-us/research/video/women-of-color-and-the-digital-labor-of-repair/> (accessed 27 November 2022).
- Nissenbaum H (2010) *Privacy in on Text: Technology, Policy, and the Integrity of Social Life*. Stanford, CA: Stanford Law Books.
- Nye DE (2003) *America as Second Creation: Technology and Narratives of New Beginnings*. Cambridge, MA: MIT Press.
- Obar JA and Oeldorf-Hirsch A (2020) The biggest lie on the internet: ignoring the privacy policies and terms of service policies of social networking services. *Information, Communication & Society* 23(1): 128–147.
- Pater JA, Kim MK, Mynatt ED, et al. (2016) Characterizations of online harassment: comparing policies across social media platforms. In: *Proceedings of the 19th international conference on supporting group work—(GROU’16)*, Sanibel Island, FL, 13–16 November, pp. 369–374. New York: ACM Press.
- Quinn K, Epstein D and Moon B (2019) We care about different things: non-elite conceptualizations of social media privacy. *Social Media + Society*. Epub ahead of print 4 September. DOI: 10.1177/2056305119866008.
- Redmiles EM, Bodford J and Blackwell L (2019) “I just want to feel safe”: a diary study of safety perceptions on social media. *Proceedings of the International AAAI Conference on Web and Social Media* 13: 405–416.
- Reynolds C and Hallinan B (2020) The haunting of GeoCities and the politics of access control on the early Web. *New Media & Society* 23: 3268–3289.
- Rheingold H (2000) *The Virtual Community: Homesteading on the Electronic Frontier*. Rev ed. Cambridge, MA: MIT Press.
- Roberts ST (2019) *Behind the Screen: Content Moderation in the Shadows of Social Media*. London; New Haven, CT: Yale University Press.

- Schwartz SH (2012) An overview of the Schwartz theory of basic values. *Online Readings in Psychology and Culture* 2(1): 11.
- Spangler PT, Liu J and Hill CE (2012) Consensual qualitative research for simple qualitative data: an introduction to CQR-M. In: *Consensual qualitative research: a practical resource for investigating social science phenomena*. Washington, DC: American Psychological Association, pp. 269–283.
- Star SL and Griesemer JR (1989) Institutional ecology, 'translations' and boundary objects: amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907–39. *Social Studies of Science* 19(3): 387–420.
- Statista (2022) Social media & user-generated content. Available at: <https://www.statista.com/markets/424/topic/540/social-media-user-generated-content/> (accessed 9 December 2022).
- Stein L (2013) Policy and participation on social media: the cases of YouTube, Facebook, and Wikipedia: policy and participation on social media. *Communication, Culture & Critique* 6(3): 353–371.
- TikTok (2021a) Community guidelines. Available at: <https://www.tiktok.com/community-guidelines?lang=en> (accessed 31 October 2021).
- TikTok (2021b) Terms of service. *TikTok*. Available at: <https://www.tiktok.com/legal/terms-of-service?lang=en> (accessed 31 October 2021).
- Twitter (2021) Privacy policy. Available at: <https://twitter.com/en/privacy> (accessed 28 September 2021).
- Vaidhyanathan S (2018) *Antisocial Media: How Facebook Disconnects Us and Undermines Democracy*. New York: Oxford University Press.
- Van der Vlist FN and Helmond A (2021) How partners mediate platform power: mapping business and data partnerships in the social media ecosystem. *Big Data & Society* 8(1): 1–16.
- Van Dijck J (2013) *The Culture of Connectivity: A Critical History of Social Media*. New York: Oxford University Press.
- Van Dijck J (2020) Governing digital societies: private platforms, public values. *Computer Law & Security Review* 36: 1–4.
- Van Dijck J, Poell T, Waal M, et al. (2018) *The Platform Society: Public Values in a Connective World*. New York: Oxford University Press.
- Wong RY (2021) Tactics of soft resistance in user experience professionals' values work. *Proceedings of the ACM on Human-Computer Interaction* 5(CSCW2): 1–28.

Author biographies

Rebecca Scharlach is a PhD candidate in the Department of Communication and Journalism at the Hebrew University of Jerusalem. Her research focuses on critical platform studies. Twitter ID: @rscarlets.

Blake Hallinan is a Senior Lecturer in the Department of Communication and Journalism at the Hebrew University of Jerusalem researching how technologies such as ratings, reviews, engagement metrics, and recommendation systems shape cultural notions of value and worth. Twitter ID: @blakeplease.

Limor Shifman is a Professor at the Department of Communication and Journalism, the Hebrew University of Jerusalem. She is currently the principal investigator of the ERC-funded project DigitalValues: The Construction of Values in Digital Spheres.