



City Research Online

City, University of London Institutional Repository

Citation: Vinhas, O. & Bastos, M. T. (2023). The WEIRD Governance of Fact-Checking and the Politics of Content Moderation. *New Media and Society*, doi: 10.1177/14614448231213942

This is the accepted version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/31596/>

Link to published version: <https://doi.org/10.1177/14614448231213942>

Copyright: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

Reuse: Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

City Research Online:

<http://openaccess.city.ac.uk/>

publications@city.ac.uk

The WEIRD Governance of Fact-Checking and the Politics of Content Moderation

Accepted for publication in *New Media & Society* (preprint version: changes still possible)

Otavio Vinhas & Marco Bastos

Abstract

In this paper we chart the conflicting standards of fact-checking outside Western, Educated, Industrialized, Rich, and Democratic (WEIRD) countries that shifted their focus from holding politicians to account to acting as content moderators. We apply reflexive thematic analysis to a set of interviews with 37 fact-checking experts from 35 organizations in 27 countries to catalog the pressures they face and their struggle with tasks that are increasingly different from the journalistic values underpinning the practice. We find that fact-checkers have to balance the number of checks across each side of the partisan divide, an exercise in ‘bothsidesism’ to manage the expectations of partisan social media users; that they increasingly prioritize the checking of viral content; and that Meta’s third-party fact-checking program prevents them from holding local politicians to account. We conclude with a discussion of our findings and recommendations for content moderation outside WEIRD countries.

Introduction

Fact-checking has become central to content moderation on social platforms, with data from Duke Reporters' Lab showing that after experiencing a rapid expansion in the aftermath of the 2016 US Presidential Election, fact-checking became a global initiative present in as many as 105 countries (Stencel et al., 2022). Substantive resources were allocated by social platforms, and Western industrialized countries aimed to develop local and regional strategies to tackle mis- and disinformation, a set of policies that placed fact-checkers in the position of unwitting online content managers (EFCSN, 2022; States, 2022). Such efforts caused considerable tension with stakeholders regarding the pursuit of a common standard for fact-checking, a problem that continues to afflict the institutional mission and values of journalists and fact-checkers (Ananny, 2018). With social platforms being the lynchpin where much political deliberation occurs, fact-checking organizations have sought to adapt their practices to the affordances and norms through which social platforms distribute and amplify content (Chadwick et al., 2022). This fundamentally asymmetric relationship with social platforms compels fact-checkers to contend with trade-offs in their traditional methods and publication strategies (Belair-Gagnon et al., 2023), a source of conflict that impinges on the core democratic values of fact-checking and its commitment toward improving public reasoning (Graves et al., 2023).

Beyond Western, Educated, Industrialized, Rich, and Democratic (WEIRD) countries, fact-checking initiatives involved in fighting misinformation face conflicting standards for acting as watchdogs of politicians and enforcing content moderation. Critical approaches to mis- and disinformation studies have shown that organizations operating in non-WEIRD contexts have to cope with various social forces eroding institutional trust and driving online harassment and physical violence (Kuo and Marwick, 2021). Notwithstanding this emerging

body of work (Kumar, 2022; Lelo, 2022; Mare and Munoriyarwa, 2022), little is known about how fact-checkers outside Western countries carry out their work in contexts of data scarcity (Cheruiyot and Ferrer-Conill, 2018), limited media freedom (Balod and Hameleers, 2019), and financial constraints to sustain their operations (Ababakirov et al., 2022). Social platforms apply a one-size-fits-all approach to their community guidelines worldwide, thereby neglecting political and cultural dimensions driving falsehoods and conspiracies locally (Jiang et al., 2021). While the empirical research remains disproportionately focused on the US context (Walter et al., 2019), the conditions described above place fact-checkers in non-WEIRD countries at the cutting-edge of strategic and pragmatic efforts to ward off untrustworthy information online.

In this article, we draw from interviews with 37 fact-checking experts to understand the impact of social platforms on the methods and strategies employed by fact-checkers operating outside WEIRD countries. As social platforms constitute the main landscape where contentious communication takes place, we probe how the governance of content moderation impinges on fact-checking practices in contexts where the management of information disorders is often overlooked (Madrid-Morales and Wasserman, 2022). To this end, this study seeks to probe a) the strategies developed by fact-checkers to gain and maintain credibility with social media users; b) changes implemented by fact-checkers to comply with the expectations of social media companies; c) the relationship between social platforms and non-WEIRD fact-checkers. We argue that while platform companies provide important support to fact-checking organizations in non-WEIRD countries, the asymmetric nature of their institutional relationship forces fact-checkers to implement perilous changes to their practices.

This study thus unpacks the convoluted relationship between the democratic mission of fact-checking and the content moderation bureaucracy of social platforms. While these

issues are global, reliance on social platforms as news sources is higher in non-WEIRD countries, where much of the internet access is provided by zero-rating data plans overwhelmingly tailored for Meta's applications (Nothias, 2020). Beyond the West, indeed, the Facebook platform is often indistinguishable from the online information space, as Meta has provided direct internet infrastructure to over 300 million people through its Connectivity (formerly Free Basics) program in the past decade (Meta, 2021). This not only renders online mis- and disinformation more effective beyond the WEIRD world (Tandoc, 2022); it also maximizes the political ramifications of content regulation for local citizens and policymakers (Gillespie et al., 2020).

In light of the above, we investigate asymmetric tensions compelling fact-checkers to work as de facto content moderators. Our findings detail the various forms through which platforms impinge on the scope, values, and institutional mission of non-WEIRD fact-checkers. To this end, our work addresses the following research questions: RQ1) What feedback do non-WEIRD fact-checkers receive from social media users? RQ2) How have non-WEIRD fact-checkers adapted their practices to meet the expectations of social platforms? RQ3) What is the relationship between social platforms and efforts to ward off mis- and disinformation in non-WEIRD contexts?

Previous Work

The shifting practice of fact-checking

The practice of fact-checking has undergone substantial changes of recent. Originally devised as a tool to verify public discourse and hold politicians to account, most fact-checking initiatives were modeled after US political newsrooms and emphasized journalism's long-established watchdog role (Graves, 2016). This traditional view of fact-checking changed

considerably in the aftermath of the 2016 US Presidential Election. In addition to correcting politician's speeches and using digital tools to improve journalistic methods, fact-checking initiatives worldwide are increasingly broadening their remit to perform content moderation on social platforms (Vinhas and Bastos, 2022). Fact-checkers thus increasingly perform content moderation work for social platforms as third-party collaborators (Gillespie, 2022), while also engaging in media literacy projects (Çömlekçi, 2022) and building knowledge databases (Nissen et al., 2022).

Literature on fact-checking provides an account of it as a practice inherited from grassroots journalism (Amazeen, 2018). Indeed, studies dedicated to measuring political and psychological effects of fact-checking found that it can reduce personal belief in misinformation across various geographical contexts (Porter and Wood, 2021; Carnahan and Bergan, 2022). However, an extensive body of work contends that fact-checking efficacy is often eclipsed by partisan motivated reasoning (Jennings and Stroud, 2021; Lyons et al., 2020), with limited potential to modify political attitudes (Nyhan et al., 2019).

Recent studies have also found fact-checking to be more effective when moderating political discourses rather than debunking posts on social platforms (Lim, 2018). In social platforms, fact-checking may unwittingly increase the overall distrust in the news (Carson et al., 2022) and is rarely circulated among groups that spread disinformation (Recuero et al., 2022). Despite these mixed results, most studies acknowledge that fact-checking increases transparency and trust in information environments (Walter et al., 2019).

The recent development of this industry is thereby strictly linked to social platforms that outsource to fact-checkers the work of labeling suspicious posts and user accounts. The most significant initiative is Meta's Third-Party Fact-Checking program (henceforth '3PFC'), where fact-checkers review flagged posts to reduce the reach of the content based on reported

levels of accuracy (Meta, 2022). Platforms such as YouTube and TikTok routinely approach fact-checkers to put together taskforces to reduce the chance of election-related interference (TikTok, 2022; YouTube, 2020), whereas Twitter appears more reluctant to work with independent fact-checking organizations (Twitter, 2022). As fact-checking organizations join in the structure of content governance created by social platforms, they have to contend with important trade-offs from these relationships (Belair-Gagnon et al., 2023).

The governance of fact-checking by social platforms

The tension between social platforms and the work of fact-checkers is to an extent unavoidable given the public mission of fact-checking and the business model of social platforms driven by advertising revenue and market share (Nielsen and Ganter, 2017). Fact-checking organizations were created by journalists, academics, and grassroots activists as part of a democracy-building movement driven by a mission to promote information transparency and political accountability (Amazeen, 2020). These early developments are somewhat at odds with the rapid adoption of fact-checking by tech companies as a central component of content moderation, whereby content-related responsibilities are outsourced to third-party actors while maintaining hierarchical and unaccountable structures of content governance (Bell, 2019; Medzini, 2021).

After investigating Facebook's partnerships with journalists and fact-checkers in the US, Ananny (2018) concluded that the framework in which fact-checkers operate was substantially molded by the asymmetrical and opaque nature of their relationships. Observers have also noted that the inclusion of independent fact-checking organizations to Meta's 3PFC influences organizations to accelerate the production of fact-checks to meet partnership targets faster (Belair-Gagnon et al., 2023). Similarly, Nissen et al. (2022) showed that social platforms employ hidden processes that determine the work of fact-checkers and shape how problematic content is selected and categorized. These developments represent a departure

from the fact-checking original mission of advocating information transparency, holding politicians to account, and enforcing journalistic principles of fairness and fact-based objectivity (Laughlin, 2022).

Content distribution on social platforms has been extensively discussed in the literature (Nieborg and Poell, 2018; Van Dijck et al., 2018), particularly platform changes that forced news professionals to trade-off editorial autonomy to optimize audience reach. But relatively little is known about how the governance of social platforms impinges on the work of fact-checkers, particularly on their methods and how they interact with audiences. Recent studies have shown that social platforms change the practice of journalism in several ways, including publication criteria, writing style, and timeline for posting news (Lischka, 2018). These changes can lead to an erosion of traditional norms that are central to building trust among audiences, ultimately undermining the authority of journalism (Ross Arguedas et al., 2022).

The affordances of platforms also play a role in mediating user interaction with content and, consequently, controlling the reach of fact-checking articles (Theocharis et al., 2022). Cotter et al. (2022) argue that the approach of social platforms to content moderation is designed to embody a libertarian vision of truth, one which assigns to users, not fact-checkers or other authoritative actors, the role of arbiters of what information should be trusted. There is also evidence showing that platforms reward affective orientations toward political news, especially negative posts containing identity-performative motivations (Chadwick et al., 2022), a development in which information is detached from institutional value (DeCook and Forestal, 2022). As a result, fact-checking is compelled to comply with practices of content amplification set out by social platforms that may conflict with the effective removal of mis- and disinformation (Petre, 2021), a process that ultimately compounds polarization and the selective reliance on fact-checks (Shin and Thorson, 2017).

The standardization of fact-checking

The association between the self-purported democratic mission of fact-checking and social platforms' dependence on independent content moderation was central to the upsurge of fact-checking, particularly outside the WEIRD context. While most initiatives in the US and Western Europe emerged as an extension of traditional newsrooms, in non-WEIRD countries fact-checkers usually operate following the 'NGO model' (Graves and Cherubini, 2016). Compared with their WEIRD counterparts, non-WEIRD initiatives have consistently struggled to achieve long-term viability while financially supporting their operations on a full-time basis. Ababakirov et al. (2022) found that most fact-checkers in Africa, Asia, and Latin America not only lack the human and technical resources to implement fact-checking in their local languages, but also contend with political pressures from local powers. Even in large media markets such as Brazil, India, and Nigeria, fact-checkers are keenly aware of their overreliance on tech companies (Lelo, 2022) and pragmatically approach such partnerships as a way to leverage their operations to other sources of funding (Nielsen and Cherubini, 2022).

Fact-checking practices are standardized by the International Fact-checking Network (IFCN), a US-based association run by the Poynter Institute that connects fact-checkers with funding partners, including social platforms. The IFCN was pivotal in extricating professional fact-checking from activism by establishing standards of transparency and nonpartisanship among practitioners, a code of practice that influenced organizations beyond their body of signatories (Rodríguez-Pérez et al., 2022; Kumar, 2022). The IFCN is also a funding proxy controlled by tech companies that liaises with news practitioners (Papaevangelou, 2023). In addition to the many contending forces shaping the standardization of fact-checking worldwide, there are localized tensions in the

implementation of fact-checking practices driven by the governance of content moderation in social platforms outside WEIRD contexts.

Data and Methods

We draw on a qualitative approach of reflexive thematic analysis (Braun and Clarke, 2022) to catalog the many tasks shaping fact-checking worldwide. Our data includes 37 semi-structured, in-depth interviews with non-WEIRD fact-checkers from 35 organizations operating in 27 countries in Africa, Asia, Latin America, and Eastern Europe (see Table 1). We included experts working in various capacities in their respective organizations (e.g., fact-checkers, editors, directors, and founders). Interviews ran between 30 and 90 minutes and were conducted in English, Portuguese, and Spanish during March to November 2021 via Zoom, except for one instance where Zoom access was restricted in the country of the interviewee and therefore WhatsApp video call was used. All interviews were transcribed and those conducted in languages other than English were subsequently translated. On two occasions, we interviewed more than one employee from the same organization and therefore the questions were tailored to their area of expertise. Most interviewees were bound by nondisclosure agreements (NDAs) with social media companies that prevented them from revealing confidential information about their partnerships. Considering the sensitivity of the topics discussed and the relatively low number of fact-checking organizations in certain continents, we opted to anonymize the identity of our interviewees.

Interviewees were recruited following a mixed approach combining purpose and snowball sampling. We aimed to build a sample that included the many different fact-checking organizations from political and cultural contexts beyond WEIRD countries. This was achieved by supplementing purpose recruiting with snowballing, which broadened the

range of interviewees and increased the geographical coverage of our sample. We started with the Duke Reporters' Lab global fact-checking database to recruit fact-checking organizations from non-WEIRD countries (Stencel et al., 2022), and subsequently snowballed to organizations that matched the inclusion criteria. From this list, we considered as legitimate fact-checking organizations those that met one of the following criteria: 1) listed as an active organization on Duke Reporters' Lab fact-checking database; 2) current signatories of the International Fact-Checking Network (IFCN); 3) declared a consistent commitment to editorial non-partisanship and financial independence on their websites, along with a detailed description of transparent fact-checking methodologies. Our recruiting process ceased once we reached saturation in the geographical area and organization type.

The resulting interviewee cohort comprises fact-checking experts from different national and linguistic contexts associated with organizations that perform various tasks related to fact-checking. Most organizations in our sample are approved signatories of the IFCN Code of Principles ($n = 27$), but our sample also covers non-signatories ($n = 10$). Notably, the majority of IFCN signatories maintain interinstitutional partnerships with social media companies, including 22 organizations that also operate under Meta's 3PFC program. While most interviewees focused on both traditional forms of fact-checking (i.e., holding politicians to account and countering political propaganda) and social media debunking practices (Graves et al., 2023), a small subset displayed a particular focus on verifying viral claims, particularly in Asia.

INSERT TABLE 1 HERE

Our interview protocol was designed to identify the various social, linguistic, cultural, and political backgrounds influencing fact-checking standards in non-WEIRD countries. The set of interview questions thus focused on three topics: 1) the moderation of social media users

by fact-checkers; 2) the shift toward moderating viral content; 3) the central role of Meta's 3PFC program. Interview data was processed by developing coding themes following iterative steps involving deductive and inductive coding (Saldaña, 2016). In the first coding cycle, we generated codes based on the semantic level of the interview responses that accounted for salient and recurring topics. In the second coding cycle, we employed axial coding to interpret our codes according to the property (context, interactions, and consequences) of fact-checkers' relationships with the social platforms they work for. Finally, we constructed our themes by grouping the resulting subthemes with the topics addressed by our research questions.

Results

RQ1: Polarized audiences

Our first theme identified that fact-checkers often adjust their work to manage polarized audiences and to safeguard their credibility, a task that often involves 'bothsiding' their social media posts to navigate the partisan fault lines. Most interviewees reported constantly receiving accusations of fabricating factual assessments to place a party or candidate in relative advantage or disadvantage. For this reason, interviewees identified partisan polarization as the main driver of mis- and disinformation in the national contexts where they work and rated polarization as the central challenge in reaching populations and groups skeptical of actors that self-identified as neutral in the ecosystem of social platforms.

The seclusion from opinion and reporting that precludes the objective evaluation of factual claims is a fundamental tenet of fact-checking. As the editor of a Latin-American organization explained, their purpose remains to supply the public debate with transparent

and impersonal assessments of information, so that “our final texts or journalistic products can be sufficient in themselves, regardless of who publishes them.” However, the editor contends that their professional commitment does not escape the polarizing grievances permeating social platforms: “if a polarized context affects you (...), it doesn’t matter the assessment we’re doing, what data or evidence we’ve presented, but who we are”. Under these circumstances, the feedback fact-checking posts receive on social platforms may bear little relation to the accuracy of the work, reflecting instead the partisan divide. The editor of an Asian organization posits that instead of moderating the public debate, fact-checkers committed to non-partisanship on social platforms must deal with emotional—and potentially violent—responses to their work, a difficult situation given their mission of promoting common ground and factual reasoning:

When you fact-check party A, party B is happy. When you fact-check that B party which was happy earlier, now they are going to target you. So it’s like everybody loves a fact-check until it is about themselves, until you touch them. So yeah, I mean, I think in general it’s about the vast amount of hate that comes our way. You’re supposed to have a thick skin, but I think it takes a toll on you eventually because you’re literally dealing with it day in and day out.

Fact-checked posts tend to reveal existing social frictions on social platforms, an issue particularly salient in non-WEIRD countries with clear religious and ethnic fault lines that are often compounded by military conflicts. One African fact-checker highlighted that, in addition to concerns about their credibility, they are often harassed and doxxed by social media mobs: “Whether they are pro-government or anti-government activists, whenever they feel that you are not serving their interests, they will insult you, and then they will publish private or false information about you.” Even non-political fact-checks can be contentious.

The director of an Asian organization recalled an incident when they “clarified that the photo circulating online was not related to the explosion that occurred in the city earlier that day, but readers started swearing and saying that we were covering up for the government.” In these contexts, social platforms are de facto information battlefields, and fact-checkers are front-line defenders taking the blunt of the casualties.

Interviewees were skeptical of building common ground on social platforms. But the editor of an Asian organization explained that working in polarizing contexts is part of their mission: “this is the main opportunity for us to make something good out of this situation, to depolarize society (...) We’re trying to connect our society again as a whole.” But interviewees were concerned that their work is only shared with a limited number of users on social platforms, and usually not the ones in most need of fact-checking content. The director of a Latin-American organization puts it thusly:

We are somewhat aware that fact-checking is something that preaches to the converted. This is the greatest challenge in what we do. Because we’ll always use a methodology that follows journalistic values. We’re not here to just move the needle to the right or to the left. Our organization checks all political groups. We apply the same news value, the same analytical values, and the same methodology to everyone we check. But then, there are bubbles that consume our fact-checks more than others. (...) The fact that I’m objective is not directly related to how people will consume my fact-check.

Interviewees listed a range of strategies to build and maintain credibility across political divides. The editor of an African organization argued that the best remedy is often to double

down on their objectivity. For a fact-checker in Eastern Europe, objectivity is achieved by demarcating political boundaries between facts and propaganda:

It is very hard to be objective because we are nationals from our country.

(...) We don't try to whitewash our politics. In this hard situation, we try to be independent. For example, we have people who support our current president and we also have pro-Russian backers. We debunk information about pro-Russian politics, but we take in their comments and try to show their positions, so we try to be independent. But, of course, if you see our material, we don't separate anti-Ukraine narratives as pro- or against Russia because they repeat each other.

Other interviewees, however, mentioned being cognizant not only of who consumes their content, but also which side will ultimately weaponize their fact-checks. Concerned that most users cannot read their work free of political biases, the director of an Asian organization stated that they try to balance their sense of objectivity if it helps clarify to audiences their neutrality: "Sometimes that would force us to reconsider it, unfortunately (...) When we are fact-checking something we shouldn't be thinking how it's going to be perceived." Another Asian director emphasized that fact-checking the two sides of the political spectrum was key to gaining credibility online: "they start to trust you when they realize that the fact-checker is addressing both sides." Such situations force fact-checkers to rethink their strict commitment to fairness and neutrality, as avoiding the weaponization of fact-checks by partisan groups may become a priority. The director of a Latin-American organization mentioned a particular strategy to safeguard their credibility:

We always highlight that it does not matter whether the fact-checker is self-employed or independent, they must also appear to be so. The public

perception of a fact-checker's legitimacy is key to their impact. Therefore, we adopt different strategies. For example, if we check a government official and we label what was said as false, then three days later we check another official and label what they said as false too. We will probably publish in our social networks a fact-check of some time ago about the leader of the opposition. Why? Because if there is a new user in that week that sees we only labeled the government or the opposition as false, this user will believe that we either work for the government or the opposition.

These strategies show the extent to which fact-checkers must balance their social media presence across the partisan divide. Strategies to manage audience perceptions may prove key to protecting the credibility of fact-checking organizations, as fact-checks often lead to identity-driven feedback on social media. While seemingly necessary in this platformized landscape, the 'bothsiding' of fact-checking do add a discretionary layer to the objectivity of fact-checks.

RQ2: The shift toward viral content

Our second research question (RQ2) unpacks the adjustments made by non-WEIRD fact-checkers to meet the expectations of social platforms. This theme highlights the emphasis on viral content as the defining parameter in selecting claims to verify, a development broadly understood as necessary to offset the speed with which falsehoods spread online.

Interviewees acknowledged that social platforms are effective channels for the replication and distribution of unverified, untrustworthy information at scale and speed. This problem was described as consequential for the escalation of cultural and political conflicts, which may often lead to physical consequences in the non-WEIRD world. Given the extent to which this problem deviates from traditional fact-checking methods, much of their work revolves around devising novel practices to offset the amplification of falsehoods on social platforms.

Interviewees described a state of permanent competition with the producers and spreaders of mis- and disinformation. The editor of an African organization emphasized it was challenging for non-WEIRD fact-checkers to avoid being buried by untrustworthy content. Similarly, the director of an organization in Asia appreciates the positive effect of the organization they work for but contends that “no matter how hard you try, the spread of false information is always faster than the truth.” With the monitoring and correcting of every piece of false information circulating online deemed impractical, fact-checkers rely on virality as a benchmark to prioritize content for verification.

The time constraint imposed by viral content is particularly challenging to non-WEIRD fact-checkers where resources are scarce or inaccessible. A Latin-American editor explained that even when they can access the official data they have to consult with experts to check if the data is reliable: “little access to public data gives us the possibility of having many interpretations of the same data, which is not ideal for fact-checking.” The editor of an African organization described the daily pressure of keeping up with viral content while depending on the responses of government officials to perform fact-checks:

Getting sources to respond is quite difficult, and with the nature of fact-checking, you have to actually work on the fact-check before the post goes any more viral than it already is. So when we’re picking our fact-checks, they’re usually at the early stages of going viral, or they have already gone viral and we have been sent to fact-check. Then you have a very short time to verify, and your sources are not cooperating with data.

Virality has nonetheless become a key factor defining what claims to verify for non-WEIRD fact-checking. Indeed, it informs the potential harm of a claim as it spreads on social platforms. Selecting claims by metrics of virality also has the added benefit of not providing

airtime to otherwise incipient problematic content. As an African editor explained, fact-checkers are acutely aware of the deceptive tactics undergirding disinformation campaigns and they avoid being used to publicize specific messages. Interviewees also recognized that assessing the potential of social rumors or partisan manipulation to cause physical outrages is essential in contentious contexts where clashes between ethnic communities or religious groups may take place. Following a local incident that resulted in approximately 50 fatalities, an Asian editor introduced quantitative variables to their risk assessment of physical harm resulting from high-impact conspiracy theories on social platforms. This was particularly salient in cases of ‘communal misinformation,’ a term often employed by Asian fact-checkers to account for problematic content where particular communities are being targeted, typically religious communities, with the Covid-19 pandemic marking a milestone after which they saw “huge amounts of communal misinformation around every event.”

On the other hand, by prioritizing viral content fact-checkers may leave other false claims unchallenged, a dilemma on which interviewees could not agree. The director of an Asian initiative contends that even if political propaganda is not viral on social platforms, it is vital to check it to provide audiences with a critical assessment of what is happening: “although people are not paying attention, it’s information that’s being repeated, maybe under the radar of the public, and if this thing goes unchallenged, it will become a fact and accepted by the public and journalists.”

Many organizations addressed this tension by segmenting their fact-checking into separated processes. Latin-American interviewees mentioned two methods to perform fact-checking, one focused on verifying public discourse, and another designed to debunk false posts on social platforms. One such organization opted to continue verifying public discourse as a separate activity from their work in debunking false posts on social platforms. According to the editor, “one thing is to take a speech or an interview by a politician or another person

of public interest and evaluate what they said. Another thing is verifying misinformation circulating in a toxic way on the internet.” Similarly, an Asian editor adapted their workflow to combine the verification of content circulating on social platforms with the original mission of their organization, directed at holding local politicians to account and identifying foreign propaganda.

Interviewees are keenly aware that while prioritizing viral content is a strategy consistent with the rationale of fact-checking, it is also mandated by Meta’s 3PFC program. The Meta partnership is the most substantive initiative assisting fact-checkers worldwide and it requires fact-checkers to adopt a viral perspective toward mis- and disinformation. In addition to being an essential source of funding (Poynter, 2022), Meta also provides fact-checkers with access to CrowdTangle, a purpose-built toolkit owned by Meta to monitor viral content that fact-checkers leverage to identify untrustworthy posts, URLs, and Facebook groups. With most interviewees prioritizing or at least incorporating virality into their workflow, the central competencies of fact-checking become largely dependent on toolkits like CrowdTangle.

Meta’s program therefore impinges on what fact-checkers define as harmful content. The stated aim of the partnership with third-party fact-checkers is to “identify, review, and rate viral misinformation” across their services (Meta, 2022). Before entering Meta’s program, an Eastern European fact-checker mainly focused on rebutting foreign propaganda, a task that requires a very specific set of competencies and contextual knowledge. But to meet the demands of the platform, they had to engage with a host of problematic content typical of social platforms: “since we started working with Facebook, we’ve been checking many conspiracy theories about vaccination.” Similarly, the director of an Asian organization, whose partnership with Meta was recently discontinued, explained that the type of content usually sent for verification by Meta was psychologically taxing on their staff and

not particularly relevant to their readership: “we had a cooperation with Facebook to fact-check fake news on social media, and they had a negative impact on us because you have to go really deep into ugly conspiracies.”

The verification work performed under Meta’s 3PFC program occurs largely behind the scenes, but it has tangible consequences to the public-facing dimension of their fact-checking work. Some interviewees discreetly pointed out that Meta provides a ‘dashboard’ displaying an endless list of viral posts awaiting verification—content flagged either by users or detected by automated tools. An editor from an Asian organization mentioned that their main responsibility with Meta is to label whether posts should be moderated based on the guidelines, a task that may deter fact-checkers from assessing relevant information deemed as trustworthy: “Some fact-checkers use the ‘true label,’ but we are not using it because there is no need to produce an article or provide an explanation to such cases.” While understanding that such tools, including CrowdTangle, are necessary in an information ecosystem where viral content has no identifiable source, the director of a Latin-American organization recognizes that Meta’s program entails tradeoffs that overshadow the public visibility of their work as fact-checkers:

We have the duty to verify 40 misinformation items. For us, it’s money, income, and also an ethical duty, because Facebook is where we’re having problems here. (...) It’s where our responsibility is bigger, the amount of work is higher, and it’s where our work is less visible.

Interviewees also listed strategies to comply with Meta’s expectations of addressing viral posts in a timely manner. An African fact-checker explained that they simplified their fact-checking process to comply with Facebook’s expectations, an adaptation for allowing posts to be checked within two to three hours: “for Facebook posts we only seek comments in exceptional circumstances, whereas for our reports we always seek a minimum of two

comments.” The interviewee voiced a common perception that working under strict deadlines discourages fact-checkers from engaging with cases that require lengthier investigations: “there is so much stuff that we have to check that we end up just moving on to the next thing that we actually can fact-check and just leave that alone. It gets frustrating, but we have to leave it.” The editor of an Asian organization argued that Meta’s standards are unsuitable for contexts with limited public data, especially when the claim is politically contentious:

Without having credible resources, it’s impossible to debunk, especially when you are labeling this on Facebook. And sometimes some facts are correct, but it’s more about malinformation [factual information decontextualized to inflict harm] (...) For instance, immigrants are usually portrayed as criminals, sexual offenders... the individual case might be right, but in the given context, when it’s part of a smear campaign designed to create the hostile attitude toward the Muslim people, it’s difficult to approach this task as debunking, so we use more analytical articles to explain context and intention.

RQ3: Meta’s 3PFC program

Our third and last research question (RQ3) unpacks the relationship between social platforms and efforts to address mis- and disinformation in non-WEIRD contexts. This theme foregrounds the asymmetric relationship between fact-checkers and social media companies in the enforcement of content moderation guidelines. Interviewees were nonetheless optimistic about Meta’s 3PFC program. The editor of an organization in Africa commented that flagging posts for Facebook has tilted the balance against bad actors, who are keenly aware that someone is out there watching. Given the centrality of platforms to the fact-checking industry, it is imperative for these organizations to establish a successful cooperation with social platforms. While some interviewees praised support provided by

social media companies to their organizations, particularly with regard to funding, others criticized the asymmetric and opaque nature of these relationships.

A recurrent problem mentioned by the interviewee cohort is that social platforms are often unwilling to enforce their own community guidelines across different national and linguistic contexts, as social platforms may prevent fact-checkers from labeling posts by local politicians—the very mission of fact-checking prior to their collaboration with social platforms. This is particularly salient in contexts of incipient democratic development, as content moderation policies that are commonly applied in WEIRD countries are often ignored in non-WEIRD contexts. Interviewees expressed concerns that the uneven enforcement of content moderation policies by social platforms makes it difficult to ward off falsehoods coming from local elite actors and leaves a plethora of requests for content removal unaddressed.

A fact-checker from an Asian organization not affiliated with Meta's 3PFC described their work routine: even when the social platforms have a flagging policy in place, enforcement is carried out by foreign fact-checkers who lack familiarity with the local context and language to assess the claims accurately. This problem is compounded in countries ruled by authoritarian governments, where fact-checkers are often perceived as censors associated with government repression. The editor of an Asian initiative mentioned that Instagram was first perceived as 'pro-government,' and then criticized due to its fact-checking system. It does not help that "there's information saying that this fact-check was done by an organization from another country." Similarly, an African editor mentioned that social platforms refuse to apply their own policies notwithstanding multiple requests for content removal: "you have people complaining about Facebook not removing some content and we just keep watching hate speech and misinformation and disinformation being spread on the platform."

Some interviewees were more skeptical of the partnership. A Latin-American director mentioned that the exercise of freedom of speech entails tangible profits for social platforms, which raises questions about the intention behind their fact-checking partnerships: “it doesn’t solve the problem because this problem is at the foundation of these platforms, of their business model. I think it’s mostly PR to be associated with the fight against disinformation than a real fight.” An African director argued that governments and policymakers in the West have been holding social platforms to account, but in African countries governments and the civil society have much less leverage. This results in a context where fact-checkers are left powerless to offset the negative impacts of mis- and disinformation locally:

We saw the efforts that went in dealing with misinformation in the US ahead of the 2020 elections. Is the same happening in the African continent when we have elections? (...) I think that there is a lot they can do to deal with misinformation and disinformation in their platforms. We are talking about very young democracies here (...) Fault lines based on ethnicity and tribal sentiments can easily be weaponized and used [against] our countries. So there’s a serious issue here, and it has to be taken seriously, because it could easily be used to undermine countries and the stability of African governments in particular.

A central problem in non-WEIRD countries is that politicians are exempted from labeling by fact-checkers according to Meta’s 3PFC program rules. These protections conferred by Meta are grounded in the understanding that political speech is the most scrutinized speech in mature democracies with a free press, a development that is justifiable, if debatable, in WEIRD countries, but that is particularly damaging for young and fragile democracies where distortions and outright lies voiced by politicians further imperils a political orders often assumed to be illegitimate. While fact-checkers have long advocated against this policy

(Mantas, 2021), interviewees emphasized the extent of the damage resulting from allowing politicians to spread disinformation. An African fact-checker cited an example where they debunked a Covid-19 conspiracy theory claiming that 5G towers recently installed in their country were responsible for transmitting the coronavirus, only to see a prominent politician rehash the same falsehood on Facebook a few weeks later, while also encouraging supporters to spread it in other channels.

This is very problematic because politicians have a big platform to spread that sort of xenophobic hatred. I think these are the most serious problems I face in this role. The fact that we can't actually fact-check politicians is a big problem because politicians have very big platforms, and their supporters tend to sort of blindly believe whatever they say, and if they can't be held accountable then our job becomes very difficult.

By preventing fact-checkers from regulating content posted by elite actors, social platforms give politicians a free pass to routinely use disinformation tactics. The director of a Latin-American organization believes politicians are keenly aware they will not face consequences for spreading falsehoods on social platforms: "their lies continue to circulate because they have many followers, great impact, and more reach than us. Perhaps for this reason they do not even deny it." The director of an organization in Asia believes fact-checkers should be granted some degree of authority to penalize the profile of bad actors instead of repeatedly warding off politicians' posts: "Facebook only takes down claims, but they don't do enough to fight the hoax makers." The rules undergirding the partnerships prevent fact-checkers from performing a core component of their mission, which is to hold politicians to account.

The common denominator across our interview cohort is that fact-checkers are not included as stakeholders despite their central work in content moderation. The partnership is

therefore not particularly transparent with regard to content-removal measures. As a Latin-American director explained, this asymmetric relationship allows Facebook to scapegoat fact-checkers when political pressure is placed on the company, while also keeping fact-checkers out of the decision-making process of content removal: “we know that when we inform Facebook that a post is false, there’ll be a reduction in reach. But they are not very transparent about how that happens.” As such, non-WEIRD fact-checkers are reduced to applying pre-established content moderation rules defined unilaterally, opaquely, and more importantly elsewhere.

Meta’s 3PFC program states that posts should be verified through “original reporting, including interviewing primary sources, consulting public data, and conducting analyses of media, including photos and videos” (Meta, 2023). For our interviewees, however, the list of verification sources recognized by Meta is incompatible with the resources available in most non-WEIRD contexts. Unfortunately, non-WEIRD fact-checkers have no leverage in forcing the platforms to develop specific community guidelines, or to prevent local governments from passing legislation that curbs or regulates speech online. Thus, by merely outsourcing content moderation tasks to local fact-checking organizations, social platforms fail to acknowledge that the development of falsehoods is intrinsically context dependent. The editor of an Asian fact-checker puts it thusly:

I feel that these policies tech giants come up with are ambiguous and arbitrary. You need to be a lot more transparent; you need to work with people who are actually working in this area. But the community at large who is actually fighting misinformation, you know, are they stakeholders? Are they part of the conversation? Should they be, maybe, because we are the other ones doing it, day in and day out?

Discussion and Conclusion

This study charted how non-WEIRD fact-checkers are adapting their workflow to address the drawbacks of cooperating with social platforms. Firstly, we discussed how non-WEIRD fact-checking organizations have to contend with accusations of bias and user harassment on social platforms. While criticism to the work of fact-checkers is to be expected (Graves, 2017), the affordances of social platforms reward affective and identity-driven engagement with information (Chadwick et al., 2022) that run counter to factual reasoning. As such, rather than offering opportunity for common ground, successful cross-cutting fact-checking posts are likely to fuel affective polarization between opposing groups (Tornberg, 2022) and be weaponized through the selective and partisan sharing of fact-checks (Shin and Thorson, 2017). While objectivity as it has been historically practiced in Western newsrooms is increasingly perceived as incongruous with journalism's pursuit of the truth (Downie Jr. and Heyward, 2023), non-WEIRD have to contend with 'bothsidesism' pressures to balance the number of checks across each side of the partisan divide, thereby trading off their objectivity to manage the perceptions of social media users.

Secondly, we unpacked how fact-checkers are increasingly prioritizing the verification of viral mis- and disinformation over other forms of harmful and untrustworthy information. Given the time-consuming nature of fact-checking, practitioners incorporate virality as the main criteria to select content for moderation. They also incorporate computational tools (e.g., CrowdTangle) to meet the endless competition for visibility on social platforms, an important departure from fact-checking's longstanding democratic mission of ensuring a well-informed public (Cotter et al., 2022). Such toolkits embody Meta's vision of problematic content, reducing mis- and disinformation to user behavior and

detaching it from the governance of content moderation (Petre et al., 2019). The constant need to address a large volume of viral posts places significant time constraints on fact-checkers, relegating them to a monitoring role (Steensen et al., 2023). This role proves challenging when it comes to dealing with politically complex claims that emerge in non-WEIRD contexts, where the availability of reliable sources is limited. Ultimately, the shift toward viral mis- and disinformation turn fact-checkers into another cog in the content moderation bureaucracy of social platforms.

Thirdly, the cooperation between fact-checking organizations and social platforms epitomized by Meta's 3PFC is an essential source of funding, but it has inadvertently moved fact-checkers away from their original responsibility of holding politicians to account. As shown by Graves et al. (2023), Meta's program prevents fact-checkers from labeling politicians on Facebook, an approach that is not only at odds with the traditional public reason model of fact-checking, but runs counter to evidence showing that falsehoods and conspiracy theories usually come from prominent public figures (Brennen et al., 2020). The purported neutrality of tech companies in non-WEIRD contexts favors local elites and amplifies their control over the local media (Cunliffe-Jones et al., 2021); in addition, it may also be conducive to social upheaval that turn social platforms into tools for violence (Sablosky, 2021). Despite the centrality of their work in content moderation and regulation of harmful speech, non-WEIRD fact-checkers are regrettably not active stakeholders of this process and are often scapegoated when concerns are raised by US and EU policymakers.

Lastly, while our findings report on the experiences of fact-checkers in non-WEIRD contexts, Western fact-checking organizations face similar threats that erode institutional trust and contend with the failure of social platforms to establish a multistakeholder framework (Douek, 2022), even if these problems remain more pronounced and widespread in non-WEIRD contexts. Further research is needed to determine whether WEIRD fact-

checkers are also being assimilated into the growing content moderation bureaucracy of social platforms.

References

- Ababakirov A, Ashour A, Bandaranayake R, et al. (2022) Meeting the Challenges of Information Disorder in the Global South. In: Wasserman H (ed). University of Cape Town.
- Amazeen MA. (2018) Practitioner perceptions: Critical junctures and the global emergence and challenges of fact-checking. *International Communication Gazette* 81: 541-561.
- Amazeen MA. (2020) Journalistic interventions: The structural factors affecting the global emergence of fact-checking. *Journalism* 21: 95-111.
- Ananny M. (2018) The partnership press: Lessons for platform-publisher collaborations as Facebook and news outlets team to fight misinformation.
- Balod HSS and Hameleers M. (2019) Fighting for truth? The role perceptions of Filipino journalists in an era of mis- and disinformation. *Journalism* 22: 2368-2385.
- Belair-Gagnon V, Larsen R, Graves L, et al. (2023) Knowledge Work in Platform Fact-Checking Partnerships. *International Journal of Communication* 17: 1169–1189.
- Bell E. (2019) The Fact-Check Industry Columbia Journalism Review.
- Braun V and Clarke V. (2022) *Thematic Analysis: A practical guide*, London: Sage.
- Brennen SJ, Simon FM, Howard PN, et al. (2020) Types, Sources, and Claims of COVID-19 Misinformation. *University of Oxford*.
- Carnahan D and Bergan DE. (2022) Correcting the Misinformed: The Effectiveness of Fact-checking Messages in Changing False Beliefs. *Political Communication* 39: 166-183.

- Carson A, Gibbons A, Martin A, et al. (2022) Does Third-Party Fact-Checking Increase Trust in News Stories? An Australian Case Study Using the “Sports Rorts” Affair. *Digital Journalism* 10: 801-822.
- Chadwick A, Vaccari C and Kaiser J. (2022) The Amplification of Exaggerated and False News on Social Media: The Roles of Platform Use, Motivations, Affect, and Ideology. *American Behavioral Scientist*.
- Cheruiyot D and Ferrer-Conill R. (2018) “Fact-Checking Africa”. *Digital Journalism* 6: 964-975.
- Çömlekçi MF. (2022) Why Do Fact-Checking Organizations Go Beyond Fact-Checking? A Leap Toward Media and Information Literacy Education. *International Journal of Communication* 16.
- Cotter K, DeCook JR and Kanthawala S. (2022) Fact-Checking the Crisis: COVID-19, Infodemics, and the Platformization of Truth. *Social Media + Society* 8.
- Cunliffe-Jones P, Diagne A, Finlay A, et al. (2021) Bad Law – Legal and Regulatory Responses to Misinformation in Sub-Saharan Africa 2016–2020. *Misinformation Policy in Sub-Saharan Africa*. 99-218.
- DeCook JR and Forestal J. (2022) Of Humans, Machines, and Extremism: The Role of Platforms in Facilitating Undemocratic Cognition. *American Behavioral Scientist*.
- Douek E. (2022) Content Moderation as Systems Thinking. *Harvard Law Review* 136.
- Downie Jr. L and Heyward A. (2023) Beyond Objectivity: Producing Trustworthy News in Today's Newsrooms. Walter Cronkite School of Journalism and Mass Communication.
- EFCSN EF-CSN. (2022) European Code of Standards for Independent Fact-Checking Organisations.

- Gillespie T. (2022) Do Not Recommend? Reduction as a Form of Content Moderation. *Social Media + Society* 8.
- 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.
- Graves L. (2016) *Deciding What's True: The Rise of Political Fact-Checking in American Journalism*, New York: Columbia University Press.
- Graves L. (2017) Anatomy of a Fact Check: Objective Practice and the Contested Epistemology of Fact Checking. *Communication, Culture & Critique* 10: 518-537.
- Graves L, Bélair-Gagnon V and Larsen R. (2023) From Public Reason to Public Health: Professional Implications of the “Debunking Turn” in the Global Fact-Checking Field. *Digital Journalism*: 1-20.
- Graves L and Cherubini F. (2016) The Rise of Fact-Checking Sites in Europe.
- Jennings J and Stroud NJ. (2021) Asymmetric adjustment: Partisanship and correcting misinformation on Facebook. *New Media & Society*.
- Jiang JA, Scheuerman MK, Fiesler C, et al. (2021) Understanding international perceptions of the severity of harmful content online. *PLoS One* 16: e0256762.
- Kumar A. (2022) Fact-Checking Methodology and its Transparency: What Indian Fact-Checking Websites Have to Say? *Journalism Practice*: 1-20.
- Kuo R and Marwick A. (2021) Critical disinformation studies: History, power, and politics. *Harvard Kennedy School Misinformation Review*.
- Laughlin AS. (2022) It's possible to be a journalist and a human. *Poynter*.
- Lelo T. (2022) The Rise of the Brazilian Fact-checking Movement: Between Economic Sustainability and Editorial Independence. *Journalism Studies* 23: 1077-1095.
- Lim C. (2018) Can FC Prevent Politicians from Lying?

- Lischka JA. (2018) Logics in social media news making: How social media editors marry the Facebook logic with journalistic standards. *Journalism* 22: 430-447.
- Lyons B, Mérola V, Reifler J, et al. (2020) How Politics Shape Views Toward Fact-Checking: Evidence from Six European Countries. *The International Journal of Press/Politics* 25: 469-492.
- Madrid-Morales D and Wasserman H. (2022) Research Methods in Comparative Disinformation Studies. In: Madrid-Morales D and Wasserman H (eds) *Disinformation in the Global South*. Hoboken, New Jersey: Wiley-Blackwell.
- Mantas H. (2021) Facebook acknowledges politicians can harm, but won't let them be fact-checked. *Poynter, Institute*.
- Mare A and Munoriyarwa A. (2022) Guardians of truth? Fact-checking the 'disinfodemic' in Southern Africa during the COVID-19 pandemic. *Journal of African Media Studies* 14: 63-79.
- Medzini R. (2021) Enhanced self-regulation: The case of Facebook's content governance. *New Media & Society* 24: 2227-2251.
- Meta. (2021) Connectivity Technologies for the Next Billion People. Meta, Inc.
- Meta. (2022) Meta's third-party fact-checking program.
- Meta. (2023) Fact-checking policies on Facebook. Meta, Inc.
- Nieborg DB and Poell T. (2018) The platformization of cultural production: Theorizing the contingent cultural commodity. *New Media & Society* 20: 4275-4292.
- Nielsen R and Cherubini F. (2022) Born in the Fire: What We Can Learn from How Digital Publishers in the Global South Approach Platforms. Reuters Institute.
- Nielsen R and Ganter SA. (2017) Dealing with digital intermediaries: A case study of the relations between publishers and platforms. *New Media & Society* 20: 1600-1617.

- Nissen IA, Walter JG, Charquero-Ballester M, et al. (2022) Digital Infrastructures of COVID-19 Misinformation: A New Conceptual and Analytical Perspective on Fact-Checking. *Digital Journalism*: 1-23.
- Nothias T. (2020) Access granted: Facebook's free basics in Africa. *Media, Culture & Society* 42: 329-348.
- Nyhan B, Porter E, Reifler J, et al. (2019) Taking Fact-Checks Literally But Not Seriously? The Effects of Journalistic Fact-Checking on Factual Beliefs and Candidate Favorability. *Political Behavior* 42: 939-960.
- Papaevangelou C. (2023) Funding Intermediaries: Google and Facebook's Strategy to Capture Journalism. *Digital Journalism*: 1-22.
- Petre C. (2021) *All the News That's Fit to Click*, Princeton, New Jersey: Princeton University Press.
- Petre C, Duffy BE and Hund E. (2019) "Gaming the System": Platform Paternalism and the Politics of Algorithmic Visibility. *Social Media + Society* 5.
- Porter E and Wood TJ. (2021) The global effectiveness of fact-checking: Evidence from simultaneous experiments in Argentina, Nigeria, South Africa, and the United Kingdom. *Proc Natl Acad Sci U S A* 118.
- Poynter I. (2022) State of the Fact-checkers 2021.
- Recuero R, Soares FB, Vinhas O, et al. (2022) Bolsonaro and the Far Right: How Disinformation About COVID-19 Circulates on Facebook in Brazil. *International Journal of Communication* 16.
- Rodríguez-Pérez C, Seibt T, Magallón-Rosa R, et al. (2022) Purposes, Principles, and Difficulties of Fact-checking in Ibero-America: Journalists' Perceptions. *Journalism Practice*: 1-19.

- Ross Arguedas AA, Badrinathan S, Mont'Alverne C, et al. (2022) "It's a Battle You Are Never Going to Win": Perspectives from Journalists in Four Countries on How Digital Media Platforms Undermine Trust in News. *Journalism Studies*: 1-20.
- Sablosky J. (2021) "Dangerous organizations: Facebook's content moderation decisions and ethnic visibility in Myanmar". *Media, Culture & Society* 43: 1017-1042.
- Saldaña J. (2016) *The Coding Manual for Qualitative Researchers*, London: SAGE Publications.
- Shin J and Thorson K. (2017) Partisan Selective Sharing: The Biased Diffusion of Fact-Checking Messages on Social Media. *Journal of Communication* 67: 233-255.
- States U. (2022) Election 2022: Fact-Checking and Countering Misinformation.
- Steensen S, Kalsnes B and Westlund O. (2023) The limits of live fact-checking: Epistemological consequences of introducing a breaking news logic to political fact-checking. *New Media & Society*.
- Stencel M, Ryan E and Luther J. (2022) Fact-checkers extend their global reach with 391 outlets, but growth has slowed. In: reports F-cc (ed). Duke Reporters' Lab.
- Tandoc EC. (2022) Contextualizing Fake News: Can Online Falsehoods Spread Fast When Internet Is Slow? In: Madrid-Morales D and Wasserman H (eds) *Disinformation in the Global South*. Hoboken, New Jersey: Wiley-Blackwell, 3-14.
- Theocharis Y, Boulianne S, Koc-Michalska K, et al. (2022) Platform affordances and political participation: how social media reshape political engagement. *West European Politics*: 1-24.
- TikTok. (2022) *Election integrity*. Available at: <https://www.tiktok.com/safety/en/election-integrity/>.
- Tornberg P. (2022) How digital media drive affective polarization through partisan sorting. *Proc Natl Acad Sci U S A* 119: e2207159119.

Twitter. (2022) *How we address misinformation on Twitter*. Available at:

<https://help.twitter.com/en/resources/addressing-misleading-info>.

Van Dijck J, Poell T and De Waal M. (2018) *The Platform Society: Public Values in a Connective World*, New York: Oxford University Press.

Vinhas O and Bastos M. (2022) Fact-Checking Misinformation: Eight Notes on Consensus Reality. *Journalism Studies* 23: 448-468.

Walter N, Cohen J, Holbert RL, et al. (2019) Fact-Checking: A Meta-Analysis of What Works and for Whom. *Political Communication* 37: 350-375.

YouTube. (2020) *Expanding fact checks on YouTube to the United States*. Available at:

<https://blog.youtube/news-and-events/expanding-fact-checks-on-youtube-to-united-states/>.

Table 1. Breakdown of the interviewed fact-checkers

No.	Position	Date	Region	Organization Type	Meta's 3PFC
1	Director	03/03/2021	Latin America	Digital News Media	Yes
2	Editor	11/05/2021	Europe / Asia	NGO	Yes
3	Director	18/05/2021	Europe / Asia	NGO	Yes
4	Fact-checker	28/05/2021	Latin America	Digital News Media	No
5	Director	04/06/2021	Asia	Digital News Media	Yes
6	Editor	04/06/2021	Asia	Digital News Media	Yes
7	Director	24/06/2021	Asia	NGO	Yes
8	Fact-checker	01/07/2021	Latin America	NGO	No
9	Editor	05/07/2021	Africa	NGO	Yes
10	Editor	06/07/2021	Europe / Asia	NGO	Yes
11	Director	09/07/2021	Asia	NGO	No
12	Director	15/07/2021	Latin America	NGO	Yes
13	Director	19/07/2021	Asia	NGO	No
14	Director	20/07/2021	Latin America	NGO	Yes
15	Editor	20/07/2021	Latin America	NGO	Yes
16	Editor	20/07/2021	Asia	NGO	Yes
17	Fact-checker	29/07/2021	Asia	Digital News Media	No
18	Editor	04/08/2021	Latin America	Digital News Media	No
19	Editor	04/08/2021	Asia	NGO	Yes
20	Editor	04/08/2021	Asia	Mainstream Media	No
21	Editor	10/08/2021	Africa	Mainstream Media	No
22	Editor	11/08/2021	Africa	NGO	No
23	Fact-checker	12/08/2021	Asia	NGO	No
24	Editor	12/08/2021	Latin America	Mainstream Media	Yes
25	Editor	14/09/2021	Latin America	Digital News Media	Yes
26	Director	14/09/2021	Latin America	NGO	No
27	Editor	15/09/2021	Africa	NGO	Yes
28	Editor	16/09/2021	Asia	NGO	No
29	Editor	17/09/2021	Africa	NGO	No
30	Director	21/09/2021	Asia	Digital News Media	Yes
31	Fact-checker	22/09/2021	Africa	NGO	No
32	Editor	04/10/2021	Asia	Digital News Media	Yes
33	Director	07/10/2021	Asia	NGO	Yes
34	Editor	16/10/2021	Asia	Digital News Media	Yes
35	Fact-checker	08/11/2021	Africa	NGO	Yes
36	Fact-checker	17/11/2021	Latin America	Mainstream Media	No
37	Fact-checker	20/11/2021	Eastern Europe	NGO	Yes