



# City Research Online

## City St George's, University of London

**Citation:** Vinhas, O. & Bastos, M. (2023). The weird governance of fact-checking: from watchdogs to content moderators. AoIR Selected Papers of Internet Research, doi: 10.5210/spir.v2023i0.13513 ISSN 2162-3317 doi: 10.5210/spir.v2023i0.13513

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

**Permanent repository link:** <https://openaccess.city.ac.uk/id/eprint/32635/>

**Link to published version:** <https://doi.org/10.5210/spir.v2023i0.13513>

**Copyright and Reuse:** Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, 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. For full details of reuse please refer to [City Research Online policy](#).



Selected Papers of #AoIR2023:  
The 24th Annual Conference of the  
Association of Internet Researchers  
Philadelphia, PA, USA / 18-21 Oct 2023

## THE WEIRD GOVERNANCE OF FACT-CHECKING: FROM WATCHDOGS TO CONTENT MODERATORS

Otávio Vinhas  
University College Dublin

Marco Bastos  
University College Dublin; City, University of London

### Introduction

Fact-checking has undergone steady transformations of recent. Originally devised as a tool designed to verify public discourse and hold politicians to account, fact-checkers have increasingly upscaled their activities to moderate viral misinformation and promote media literacy initiatives (Çömlekçi, 2022; Graves & Mantzarlis, 2020). In the aftermath of the 2016 US Presidential Election, policymakers and social platforms invested in fact-checking initiatives to restore truth and consensus reasoning in public debate (Bennett & Livingston, 2021). Substantive resources were allocated by social platforms, and Western industrialized countries developed local and regional strategies to tackle mis- and disinformation, a set of policies that placed fact-checkers in the position of unwitting online content moderators (EFCSN, 2022; States, 2022). Such efforts, however, clashed with a common standard for fact-checking and caused considerable tension with stakeholders (Bélair-Gagnon et al., 2022), a problem that continues to haunt the institutional mission and values of journalists and fact-checkers (Ananny, 2018). With social media hosting the ecosystem where much political deliberation occurs, unpacking the governance of fact-checking by social platforms is thus crucial to understanding the sociotechnical entanglements determining the digital infrastructure of knowledge and facts (Ford, 2022).

Beyond Western, Educated, Industrialized, Rich, and Democratic (WEIRD) countries, fact-checking initiatives follow inconsistent standards for acting as watchdogs of politicians and enforcing content moderation. Critical approaches to mis- and disinformation studies have shown that fact-checkers operating in non-WEIRD contexts must cope with various social forces eroding institutional trust and driving online harassment and physical violence (Kuo & Marwick, 2021). While these issues are

Suggested Citation (APA): Vinhas, O., Bastos, M. (2023, October). *The Weird Governance Of Fact-Checking: From Watchdogs To Content Moderators*. Paper presented at AoIR2023: The 24th Annual Conference of the Association of Internet Researchers. Philadelphia, PA, USA: AoIR. Retrieved from <http://spir.aoir.org>.

global, the reliance on social platforms as news sources is higher in non-WEIRD countries, where much of the internet infrastructure is restricted to Meta's applications (Nothias, 2020). Additionally, fact-checkers beyond the US and Western Europe generally depend on partnerships and grants from tech companies to maintain their operations, as many initiatives evolved from informal journalism and online activism instead of developing from a liberal and profitable media market (Ababakirov et al., 2022). These conditions not only make the eroding potential of mis- and disinformation more effective beyond the WEIRD world (Tandoc, 2022); they also maximize the political ramifications of content regulation for local citizens and policymakers (Gillespie et al., 2020).

## **Method**

We interviewed 37 fact-checkers from 35 organizations across 27 non-WEIRD countries in Africa, Asia, Latin America, and Eastern Europe. The interviews lasted between 30 and 90 minutes and were conducted in three languages (English, Spanish, and Portuguese) between March and November of 2021. We used Duke Reporters' Lab global fact-checking database as our reference point to recruit participants and subsequently snowballed to organizations that matched the following inclusion 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. We cease our recruiting process once we reached saturation across geographical areas and organization types. The interview questions revolved around three main topics: 1) social media moderation by fact-checkers; 2) the criteria implemented to identify false and harmful content; 3) the role of social platforms in regulating the public debate.

Our analysis relied on Braun and Clarke (2022) reflexive thematic analysis framework and was devised to address the following research questions: 1) How have non-WEIRD fact-checkers adapted their practices to meet the expectations of social media platforms? 2) What feedback do non-WEIRD fact-checkers receive from social media users? 3) What is the relationship between social platforms and efforts to address mis- and disinformation in non-WEIRD contexts? Iterative coding through NVivo was applied to the transcribed interviews to identify recurring topics that appeared in the data. Following the first round of semantic coding, we applied axial coding to interpret the data in line with literature on news and platformization. The processed data reveals the relationship (context, interactions, and consequences) between fact-checkers and social media platforms.

## **Discussion and Conclusion**

Our findings detail three ways through which social platforms impinge on the scope, values, and institutional mission of non-WEIRD fact-checking organizations. Figure 1 shows the themes developed in the analysis that cataloged the methods, tools, and strategies fact-checkers implemented to produce pragmatic solutions to ward off the scourge of problematic content online.

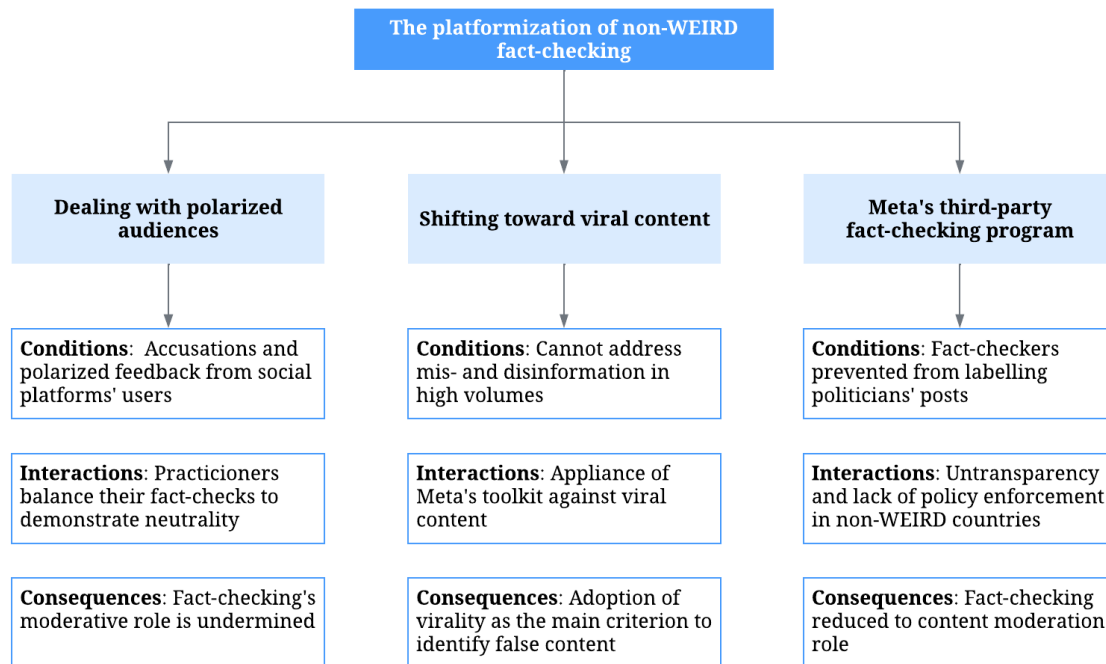


Figure 1: Conditions, interactions, and consequences of the platformization of fact-checking

Our themes shed light on how non-WEIRD fact-checkers have adapted their workflow to meet the expectations of cooperating with social media platforms. Firstly, accusations of bias and weaponization of fact-checks on social platforms compel fact-checkers to trade off their objectivity to balance fact-checking posts across partisan divides. As social platforms' affordances reward affective and identity-driven engagement with information over factual reasoning (Chadwick et al., 2022), they inherently disrupt the informative role of fact-checking organizations (Cotter et al., 2022). Secondly, to tackle mis- and disinformation in high volumes, fact-checkers are increasingly adapting their work to compete for visibility on social platforms (Petre et al., 2019). By adopting Meta's toolkit designed to verify viral content, fact-checking organizations embody Facebook's vision of problematic communication, where mis- and disinformation are reduced to a behavioral problem detached from contextual factors (Anderson, 2021). Thirdly, social platforms fail to apply their own community guidelines in non-WEIRD contexts, despite partnering with seasoned local fact-checkers speakers. Lastly, Meta's third-party fact-checking program prevents fact-checkers from labeling politicians on Facebook. This is in sharp contrast to the traditional watchdog mission of fact-checking. It is also at odds with evidence showing that falsehoods usually come from prominent public figures (Brennen et al., 2020).

Taken together, the platformization of non-WEIRD fact-checkers entails a convoluted process in which social media platforms gradually nudge fact-checkers into becoming part of the content moderation industry, a shift that runs counter to the democracy-building values underpinning the fact-checking movement (Amazeen, 2020). Indeed, by including non-WEIRD fact-checkers in their taskforces against mis- and disinformation, social platforms offer crucial funding for initiatives operating in contexts where fact-checking would rarely be sustainable. But fact-checkers have no input over the policies and day-to-day decisions determining what content should be removed or reduced, an arrangement that leaves them vulnerable to polarization and being scapegoated as censors in the platform. This is particularly the case for Meta, the most prominent funder of non-WEIRD fact-checking initiatives.

Given the above, we conclude that fact-checking in non-WEIRD countries is being deprived of its journalistic value as it is assimilated into the content moderation industry. The governance of non-WEIRD fact-checking by tech companies such as Meta and Google entails the strategic capture of third-party news practitioners (Papaevangelou, 2023) rather than the forging of asymmetrical but mutually-dependent partnerships (Belair-Gagnon et al., 2023). While our findings reflect the experiences of fact-checkers operating in non-WEIRD contexts, they nonetheless establish the groundwork for investigating their WEIRD counterparts. Western fact-checkers address similar threats that erode institutional trust and contend with the failure of social platforms to establish a multistakeholder framework to address mis- and disinformation, even if these problems remain more pronounced and widespread in non-WEIRD contexts.

## References

- Ababakirov, A., Ashour, A., Bandaranayake, R., De Silva, R., Diouf, D., Galpaya, H., . . . Valente, J. (2022). *Meeting the Challenges of Information Disorder in the Global South*. <https://idl-bnc-idrc.dspacedirect.org/handle/10625/60954>
- Amazeen, M. A. (2020). Journalistic interventions: The structural factors affecting the global emergence of fact-checking. *Journalism*, 21(1), 95-111. <https://doi.org/10.1177/1464884917730217>
- Ananny, M. (2018). *The partnership press: Lessons for platform-publisher collaborations as Facebook and news outlets team to fight misinformation*.
- Anderson, C. W. (2021). Fake News is Not a Virus: On Platforms and Their Effects. *Communication Theory*, 31(1), 42-61. <https://doi.org/10.1093/ct/qtaa008>
- Bélair-Gagnon, V., Graves, L., Kalsnes, B., Steensen, S., & Westlund, O. (2022). Considering Interinstitutional Visibilities in Combating Misinformation. *Digital Journalism*, 10(5), 669-678. <https://doi.org/10.1080/21670811.2022.2072923>

- Belair-Gagnon, V., Larsen, R., Graves, L., & Westlund, O. (2023). Knowledge Work in Platform Fact-Checking Partnerships. *International Journal of Communication*, 17, 1169–1189. <https://ijoc.org/index.php/ijoc/article/view/19851>
- Bennett, W. L., & Livingston, S. (2021). A Brief History of the Disinformation Age: Information Wars and the Decline of Institutional Authority. In W. L. Bennett & S. Livingston (Eds.), *The Disinformation Age: Politics, Technology, and Disruptive Communication in the United States*. Cambridge University Press.
- Braun, V., & Clarke, V. (2022). *Thematic Analysis: A practical guide*. Sage.
- Brennen, S. J., Simon, F. M., Howard, P. N., & Nielsen, R. K. (2020). *Types, Sources, and Claims of COVID-19 Misinformation*
- Chadwick, A., Vaccari, C., & 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*.  
<https://doi.org/10.1177/00027642221118264>
- Çömlekçi, M. F. (2022). Why Do Fact-Checking Organizations Go Beyond Fact-Checking? A Leap Toward Media and Information Literacy Education. *International Journal of Communication*, 16.  
<https://ijoc.org/index.php/ijoc/article/viewFile/19728/3903>
- Cotter, K., DeCook, J. R., & Kanthawala, S. (2022). Fact-Checking the Crisis: COVID-19, Infodemics, and the Platformization of Truth. *Social Media + Society*, 8(1).  
<https://doi.org/10.1177/20563051211069048>
- EFCSN, E. F.-C. S. N. (2022). *European Code of Standards for Independent Fact-Checking Organisations*.  
<https://eufactcheckingproject.com/app/uploads/2022/10/EU-CODE-EFCSN-.pdf>
- Ford, H. (2022). *Writing the Revolution : Wikipedia and the Survival of Facts in the Digital Age*. The MIT Press.  
<https://doi.org/https://doi.org/10.7551/mitpress/11386.001.0001>
- Gillespie, T., Aufderheide, P., Carmi, E., Gerrard, Y., Gorwa, R., Matamoros-Fernández, A., . . . Myers West, S. (2020). Expanding the debate about content moderation: Scholarly research agendas for the coming policy debates. *Internet Policy Review*, 9(4). <https://doi.org/10.14763/2020.4.1512>
- Graves, L., & Mantzarlis, A. (2020). Amid Political Spin and Online Misinformation, Fact Checking Adapts. *The Political Quarterly*, 91, 585-591.  
<https://doi.org/https://doi.org/10.1111/1467-923X.12896>

- Kuo, R., & Marwick, A. (2021). Critical disinformation studies: History, power, and politics. *Harvard Kennedy School Misinformation Review*.  
<https://doi.org/10.37016/mr-2020-76>
- Nothias, T. (2020). Access granted: Facebook's free basics in Africa. *Media, Culture & Society*, 42(3), 329-348. <https://doi.org/10.1177/0163443719890530>
- Papaevangelou, C. (2023). Funding Intermediaries: Google and Facebook's Strategy to Capture Journalism. *Digital Journalism*, 1-22.  
<https://doi.org/10.1080/21670811.2022.2155206>
- Petre, C., Duffy, B. E., & Hund, E. (2019). "Gaming the System": Platform Paternalism and the Politics of Algorithmic Visibility. *Social Media + Society*, 5(4).  
<https://doi.org/10.1177/2056305119879995>
- States, U. (2022). *Election 2022: Fact-Checking and Countering Misinformation*  
<https://www.state.gov/briefings-foreign-press-centers/fact-checking-and-countering-misinformation>
- Tandoc, E. C. (2022). Contextualizing Fake News: Can Online Falsehoods Spread Fast When Internet Is Slow? In D. Madrid-Morales & H. Wasserman (Eds.), *Disinformation in the Global South* (pp. 3-14). Wiley-Blackwell.