A Survey of Public-Oriented Organizations Analyzing Social-Media Disinformation

Darren L. Linvill & Patrick L. Warren

Media Forensics Hub

Clemson University

INTRODUCTION

A 2019 report from Oxford's Computational Propaganda Research Project suggested that more than 70 countries globally operate social media disinformation campaigns (Bradshaw & Howard). This number is impressive given that the strategy was virtually unknown less than a decade ago.

While the use of disinformation by motivated actors with political or economic agenda is far from new (MacDonald, 2017), social media offers an inexpensive and often powerful vehicle for such activity. Platforms facilitate the concealment of provenance, speed the rapid adaptation of tactics and narratives, and provide the ability to efficiently target vulnerable populations. Disinformation has threatened the integrity of democratic processes, undermined trust in science, and fostered ideological division across the globe and the struggle against this threat has been called our new "forever war" (Nunberg, 2019).

The turning point was, perhaps, the discovery of the Internet Research Agency. In 2014 the Russian Internet Research Agency (IRA) – owned by oligarch Yevgeny Prigozhin and widely held to be a tool of President Putin and the Russian state (Shane & Mazzetti, 2018) – began to sow discord in the U.S. political system. The information operation run by the IRA was done largely by engaging in coordinated inauthentic behavior on social media platforms including Facebook, Instagram, Twitter, and Reddit. The IRA used these tools to spread disinformation to unwitting audiences. Since Russia's perceived success, a range of both state and non-state actors have begun to engage in similar operations through a range of platforms.

Since 2016, a range of non-profit, private sector, and academic organizations aimed at slowing and stopping disinformation have been established. Many such organizations conduct investigations with the goal of identifying and understanding influence operations and other propagators of disinformation spread through social media. The goal of this report will be to describe the current state-of-the-art regarding the work such organizations undertake and, in the process, describe factors which limit the impact of the work these groups produce. This research will focus on those non-state organizations with a primary mission of understanding and combatting social media disinformation and who conduct research which investigates ongoing campaigns. Here we exclude organizations that exclusively conduct peer-reviewed academic work, those that are primarily engaged in fact-checking of individual suspicious claims, and those that conduct general political/technology journalism whose work sometimes includes covering these sorts of operations. Instead, we examine organizations for whom the analysis of these sorts of bad actors for public consumption is a core part of their mission. In other words, those public facing groups that serve an important new public service mission in the digital age.

More specifically, we will:

1. Identify appropriate groups which publish open-source investigations of social media influence campaigns and characterize these along a range of factors.
2. Detail the nature of the work in which such groups engage, including typical targets of investigation as well as methods of data collection and analysis.
3. Describe major limitations faced by these organizations, both stated and inferred, and make recommendations for addressing them.

METHOD

The initial phase of research required us to identify the population of organizations to be analyzed. We first established a criterion for inclusion: "Does this organization analyze coordinated inauthentic information operations or other organized bad actors on social media as a core element of their mission?" As previously stated, we specifically excluded government agencies, organizations with a journalistic mission, as well as organizations which exclusively publish peer-reviewed, academic research. Initial internet and social media searches employing various Boolean search terms resulted in 88 organizations which required further analysis.

Following criterion sampling, we developed a coding process to use with each organization (see appendix). Each organization was coded for a series of qualities. These included items such as the number of staff, sources of funding, and platforms analyzed as well as methods of data collection, analysis, and attribution. Reports published by each organization between May 1, 2021 and May 31, 2022 were also collected and each report was coded for the location of the target of the disinformation described in the report as well as the origin of the campaign (either explicit or implicit). Following initial coding, 32 organizations were retained as qualifying under our sampling criteria. The organizations removed were primarily organizations such as media literacy and fact checking organizations with missions related to our target organizations, but which did not themselves conduct research on social media disinformation campaigns. The full survey/codebook is attached as Appendix A.

Following initial coding we conducted snowball sampling to ensure that we collected as close to the full population of organizations meeting our study criteria as was reasonable. All 32 remaining organizations were contacted by email. In this email we described the nature of our research and asked the recipient to share with us the names of any organizations which should be included for further analysis. In addition, we contacted several researchers known to us personally who we felt may have insight and requested their input as well. Following snowball sampling one organization was removed from the study (as it was found to be part of a larger organization already included) and eight additional organizations were identified and coded, bringing the total number of organizations analyzed to 39. These organizations are listed as Appendix B.

RESULTS

Organizational Structure

Of the groups analyzed, 32 were organized as non-profits, with ten of these being part of a larger academic institution (e.g. Stanford Internet Observatory) and 22 not formally affiliated with an academic institution (e.g., DFRLab). Seven groups were for-profit, threat intelligence and cybersecurity firms (e.g. Mandiant). The not-for-profit groups funding came from a mix of state and private sources; for those groups that acknowledged sources of funding seventeen included state sources and 21 included private foundations. The organizations analyzed were predominately relatively small. Nineteen groups had fewer than fifteen full-time employees. Only six groups were larger than fifty employees, and half of these were for-profit organizations. For all large groups it was difficult to assess from public data the true number of staff dedicated to analyzing social media information operations given that these groups had multiple reasons for being.

The 39 organizations assessed in this report produced 234 reports exploring social media disinformation in the study period. The larger, more well-resourced organizations clearly accounted for the plurality of these reports, however. The median number of reports produced per group was three. Below we analyze first the nature and sourcing of data which appears in these reports and then the forms of analysis various groups engaged in. Finally, we will explore the nature of the campaigns being explored. Figure 1 illustrates the distribution of organization size and structure alongside the number of reports produced by each categorization.

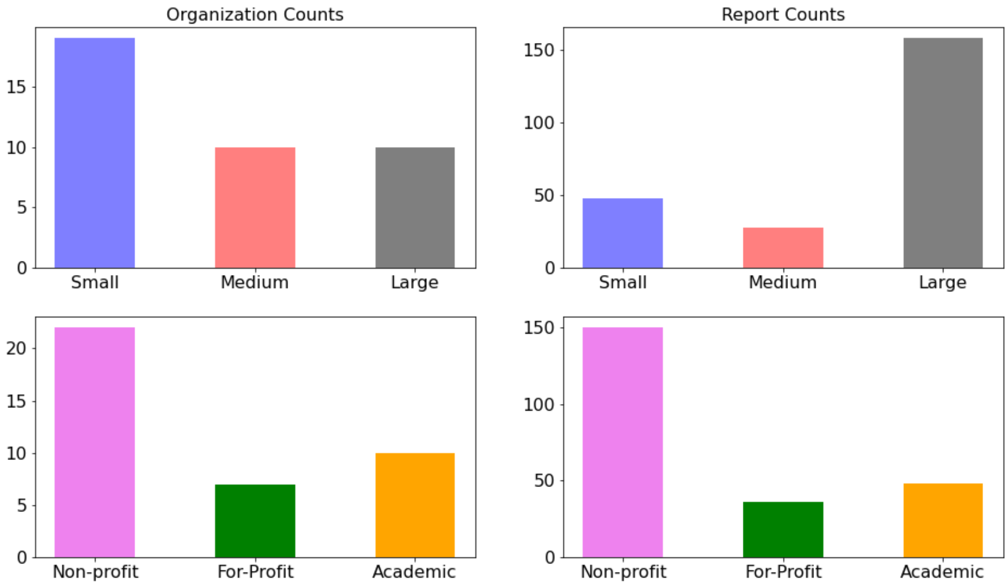
For the analysis performed below we split the 39 organizations into three distinct groups:

- **Small** (n = 19): These groups had fewer than 15 staff. Together, they produced 48 reports. 6 of these organizations actually produce no qualifying reports over the past year, despite having the public analysis of bad actors as part of their mission.
- **Medium** (n = 10): These groups had more than 15 staff but produced five or fewer reports each. Together, they produced 28 reports.
- **Large** (n = 10): These groups had more than 15 staff and produced more than five reports each. Together, they produced 158 reports.

The reports we discovered and analyzed were extremely concentrated in the "large" organizations, which despite being only a quarter of the organizations doing this work produced two thirds of the reports. This suggests, to us, that their processes are probably very different from the smaller organization and should, therefore, not be pooled together for our analysis. Since "medium" organizations had many employees but did not produce many reports, they did not seem to fit with either of the two archetypes, so we also divide them out into their own class.

Similarly, as the motivations and goals of for-profit, academic, and non-academic non-profit organizations might substantially differ, we also provide results the divide the organizations along these lines. Here, unsurprisingly, non-academic, non-profits are both the bulk of organizations and the bulk of reports.

Figure 1. Counts of Organizations and Reports by Organizational Form and Size.



Platforms and Data

Figure 2 presents the share of organizations, of each type, that conducted at least one analysis of each platform. The overwhelming majority of organizations analyzed Twitter and Facebook, and indeed most reports included one or both of these platforms. Just under half of all organizations analyzed Telegram, YouTube, and Instagram, which rounded out the top 5. Analysis becomes extremely sparse for platforms beyond these five. The platform rankings were also nearly identical across organizational form and scale, although there was a longer tail for medium and large organizations.

Aligning with the platform pattern, data acquisition practices are very concentrated. Figure 3 illustrates the share of organizations, of each type, that used each data acquisition method at least once. Most organizations acquire their data from platforms APIs. Frequent methods also included web scraping, human collection, and subscription data services. Larger, more well-resourced organizations acquired data from a more heterogeneous set of sources, relative to smaller counterparts. This was particularly true with regards to privileged access to data from the platforms. From the very small number of groups that utilized data supplied through a direct relationship with the platforms (n = 6), only one had fewer than 15 employees. The ranking of data acquisition methods was very similar across organizational scale and form. The only substantial deviation was the relative infrequency with which large organizations used web scraping, perhaps because they had much more access to alternative methods such as subscription services that provided the same data. For-profits organizations were also more likely to avail themselves of subscription data services.

Most organizations primarily analyzed the same sorts of data. Figure 4 shows the share of organizations of each type that analyzed each type of data at least once. The textual content of messages was the most consistent sort of data employed, but image data from messages and textual and image data from accounts were also common. Other relatively common forms of data employed were hashtags employed by particular operations, video content, and network links between accounts. Of note, metadata, whether of individual message or of media content, was seldom analyzed. Despite a general tendency to use a wider variety of techniques, there were few substantial differences in the ranking of data use between larger, more well-resourced organizations relative to their smaller counterparts. The only substantial deviation was in the use of video data, which was much more common among large organizations. Academic organizations used certain sorts of data very consistently-- including textual message and account data in nearly all their reports. For-profit firms stood out for eschewing data at the message level, and instead focusing on social-network and account level data, including archival data.

Analytical Techniques and Weaknesses

For all organizational types and scales, the most consistently used analytic technique was visualizations of the chronological timing of messages (see Figure 5). Other common techniques included network mapping to illustrate relationships between accounts and qualitative thematic analysis of message content. The ranking of analysis techniques was largely consistent across organizational scales, except that large organizations depended a lot more on cross-platform analysis, a bit more on human labelling, and much less on topic modelling than did small and medium organizations. Here, in most cases, topic modelling was implemented through were sets of key words and those topics were tracked through timing visualizations.

Not all organizations or all reports attributed campaigns to specific actors, but when they did we saw a range of methods employed. The most commons included the discussion of similarities in tactics between campaigns, content similarities between campaigns, acknowledging networked activity (particularly links to specific, authentic accounts), and the utilization of attributions made by platforms (see Figure 6). This last often came in the form of partnerships with platforms (Twitter and Meta specifically) where the platform would share attributed data with a non-governmental partner as a part of the platforms process of disclosure. This helps to explain the fact that this method appears far more common among large organizations.

Organizations did not frequently explicitly acknowledge limitations faced in producing their reports. When they did, however, they most frequently pointed at the nature of the platforms themselves. The most common limitation suggested was privacy settings or other structural elements inherent to social media platforms (see Figure 7). This was followed closely the “memory hole” problem, the fact that data which has been deleted or suspended could not be accessed. The rankings of limitations was similar across scale and form, but—as you might expect—academic organizations were much more likely to explicitly mention a weakness or limitation than other. One for-profit organization mention one limitation (Memory Hole) in one report.

Despite this lack of explicit acknowledgement, the actual limitations in many of the reports were substantial, at least when judged against an infeasible standard of unconstrained analysis using state-of-the-art methods. These limitations are presented in Figure 8. By far, the most common weakness is what we’ve coded as “Lack of expertise.” Analyzing the behavior of bad actors on social media is complex and multidisciplinary, with an ever-changing frontier of techniques. But most reports are nowhere near the frontier, using the simplest descriptive methods, not because those methods are the best way to address the questions at hand but rather, it appears, because the authors are unaware of better methods or do not know how to execute them. This need not be computational or statistical expertise, although it often is. It can also include social-scientific, historical, or domain expertise.

Expertise was not the only substantial limitation. The number of platforms analyzed, memory hole, and changes in behavior were also substantial problems. The ranking of the limitation was quite similar across organizational scales, although the larger organization did, in fact, suffer from fewer other limitations, despite their large number of reports. The pattern was a bit more complex across organizational forms, failing to track activities across platforms was actually the problem suffered by most academic organizations, and changes in CIO behavior was just as important as expertise. This, perhaps, indicates that the academic organizations are working a bit closer to the methodological frontier, but do not quite have the breadth to track large campaigns across platforms or time.

Geographic Focus of Reports

When examining each of the 234 reports individually, we find 48 of them explore disinformation that was inherently global in context. The remaining reports all addressed campaigns that were not general but specifically targeting users in one or more nations, regions, or cultural communities. By far the most common targets of campaigns studied were located in Europe (see Figure 9). We do see some differences in the nature of campaigns examined when we look at large, well-resourced organizations relative to their smaller counterparts. Large organizations were more likely to draft reports exploring general campaigns or trends in disinformation, campaigns or issues which did not have a specific national or regional focus. They were also proportionately less likely to explore campaigns targeting nations or regions in South America. Academic organizations are actually a little more likely to issue reports on operations that target North America, while for-profit and non-profit organizations overwhelming focus on Europe.

DISCUSSION

Overall, the degree of consistency across groups engaging in research exploring disinformation spread through social media is compelling. There is a high degree of uniformity along several dimensions: the platforms typically analyzed, the types of data employed, the forms of analysis, and the targets of research. We see more similarity then difference regardless of the size or form of the organization. This suggests that the field as a whole may be facing many of the same obstacles and limitations, independent of resources or institutional challenges.

Platforms and Data

The findings outlined above suggest several items worthy of further discussion. First, and perhaps most obvious, is the overwhelming bias we see in which social media platforms are being analyzed. All groups, regardless of size or purpose, typically focused their reporting on one or more of five platforms: Twitter, Facebook, Telegram, YouTube, and Instagram. Some bias is understandable, particularly a preference for the largest of platforms. Facebook and YouTube are the first and second largest platforms respectively and have nearly three billion active users each. Twitter (396M active users), however, was analyzed by nearly every group included in this study despite having nearly half the active users of TikTok (732M active users) which only appeared in reports from four groups. Several Chinese platforms with hundreds of millions of active users received no meaningful attention whatsoever.

Reasons for this imbalance, of course, are found partially in the data regarding how organizations collect and analyze data. First, and perhaps most important, reports are dependent on data, and most groups are in turn very dependent on data collection being facilitated through APIs, subscription data services, or the platforms directly. We know that these organizations are not focused on the same five social media platforms because there is no bad behavior to monitor elsewhere; disinformation on smaller platforms is well documented (e.g. Silverman & Kao, 2022; Timberg, 2020). All of the top five platforms have an API that can be used to access data for research purposes (even if it is not quite designed with that purpose in mind), but the top two (Twitter and Facebook) have made providing researcher-specific mechanisms through which they facilitate the collection or reception of data for analysis, with Twitter (#1) being exceptionally accessible to researchers.

While Telegram does not facilitate data collection by researchers, it does offer another compelling advantage over some other platforms that helps explain why it remains a focus. Telegram is a largely text-based platform and our analysis showed a clear bias for text as a unit for analysis. Platforms that are largely image or video-based offer larger hurdles for tracking disinformation at scale. Images and video are harder to search and, in the case of video, can take significant time to view while offering hurdles to translation from other languages. It is a known fact that text analysis remains the focus of most professional fact checkers (Nakov, et al., 2021) and it appears that this is also the case with groups tracking information operations more broadly.

Pervasive Limitations in Analysis

All organizations analyzed for this report focused on similar analysis procedures, typically using thematic analysis and visualizations of data. The more advanced reports employed network analysis, showing relationships between accounts in network maps. Statistical analyses were remarkably rare, and often limited to descriptive statistics. Reporting on ongoing disinformation operations in a descriptive and illustrative manner is without question valuable and serves an important role in civil society that we too seldom see from academia, which faces different goals and constraints, or journalists, who have limited time and resources.

Purely descriptive analyses have inherent limitations, however. Straightforward descriptions of disinformation campaigns typically fail to answer important questions regarding the relative importance of these campaigns. They offer no important comparisons or statistical understanding of the effect these campaigns may have on platform conversations or the broader media ecosystem. A social media disinformation campaign may disseminate a thousand messages from a thousand different inauthentic accounts, but simply identifying that fact and decontextualizing it from the ecosystem it is a part of can give a blinkered view of reality and a skewed understanding of the role disinformation may play in public discourse.

Working with organizations to address limitations in their analysis may be a heavy lift. The groups studied for this report offered few acknowledgements of limitations, and when they did limitations were focused on the nature of the platforms and not on their own expertise or techniques. Academic institutions were somewhat more likely to acknowledge limitations (in keeping with standards of academic writing), particularly in comparison to for-profit organizations, who addressed a research limitation only once in a single report. Many of these organizations do not seem to know what they don't know, perhaps driven by the relative nascence of the field. Social media disinformation research is less than a decade old and many analytic approaches are still themselves being developed.

Global Distribution of Analysis

Finally, the organizations included in this report disproportionately explore campaigns targeting European nations, regions, or cultural groups and campaigns originating, overtly or implicitly, from the Russian state. Perhaps the Russian invasion of Ukraine and subsequent associated disinformation campaigns targeting the West may explain some of the disparity—the issue was prevalent in the last quarter of the study period. The previously mentioned focus on a limited number of platforms may also contribute to this imbalance; platforms' popularity differs regionally and culturally. These issues alone, however, cannot explain a six to one disparity between reports addressing campaigns targeting Europe relative to reports addressing campaigns targeting Africa. This point is particularly important given the fact that disinformation targeting the global South is a well acknowledged problem. Of the eight data sets in Twitter's most recent (December 2021) public release half originate from Africa and a quarter from Latin America.

OPPORTUNITIES FOR IMPROVEMENT

The analysis presented in this report suggests several specific areas where investment has the potential to expand and improve the nature of public-facing research in social media disinformation.

1. **Broader range of data sources.** Research is currently focused on only a relatively narrow slice of the social-media ecosystem. Many major platforms were left virtually ignored among the hundreds of reports reviewed here. It seems probable that this is due to lack of access, difficulty in analyzing non-textual data, or a combination of both factors.

   Twitter, as a platform, has made several decisions which benefit researchers engaging with their platform. These efforts, combined with the fact that Twitter is a largely text-based platform and the fact that typical users do not set their accounts to private, results in a disproportionate focus of research on the platform relative to its user base. Twitter serves as a model for what hurdles need to be overcome, it is clearly offering researchers what they need. We must facilitate researchers having the same ease of engagement on a much broader range of platforms. This could mean building user friendly tools to access data, improved processes for analyzing non-textual data, and perhaps even direct engagement with platforms to negotiate disclosures.

2. **Standard definitions and methodologies** Similarly, to move forward, the current interdisciplinary study of social media disinformation requires common metrics to define the strategy and tactics employed by information operations. The meanings of many common concepts addressed in the study of disinformation are all too often assumed and in practice implemented differently across groups and reports. Common operationalizations would help to regulate the field as a whole, and the work of public facing researchers specifically. It would also lower the barriers to contributing, as applying standardized methods requires less specialized knowledge and might allow organizations with regional expertise but lacking methodological expertise to contribute.

3. **Statistical/behavioral norms.** One driver, we think, of the lack of context provided in most reports is the lack of easily available norms for organic and inorganic behavior. Even the best description of what a problematic group does is hard to interpret without reference to these sorts of norms. It is, of course, not feasible for a single organization to develop these norms, as they are interested in and have a window on only a very small part of the social- media conversation. And even an aggregation from individual studies will be difficult without standard methodologies (See opportunity 2) and a sufficiently broad view (see opportunity 4).

4. **A global perspective.** Social media does not have national boundaries and the impacts of disinformation are never constrained to geographic barriers. It is crucial to understand disinformation globally, and to do so we must encourage researchers to look past the political conversations that may impact them most directly and engage in research in cultures and contexts that may not be immediately familiar to them. We know that disinformation disproportionately impacts autocratic nations with fewer democratic freedoms (Linvill & Warren, 2021) and so social media conversations impacting these places must be a focus of future effort. The benefits will not be restricted to those places but would themselves be global.

RECOMMENDATIONS

To achieve this progress on exploiting these opportunities, we recommend taking the following steps:

1. **Develop a cross-cutting research platform than spans as much of social media as feasible.** Negotiate and/or engineer API-like access to myriad platforms under a standard framework, to include both flows of regular activity and activity that has been marked for removed from the platform, with appropriate protections for each. This might require some hardball politics and guarantees of legal and privacy protections. CrowdTangle and the Twitter API/IO Archive are a place to start, but they still suffer from significant memory-hole problems. Especially important is to maintain a more complete record of the behavior of the suspended accounts than is currently available in any platform release, including their networks, interactions, and behavior over time.
2. **Create and maintain tools that allow the easy implementation of cutting-edge analytical techniques on data from this research platform.** These tools should be well documented, constantly updated, and they will become the de-facto industry standard. Good examples include, for example, topic modelling, network analysis, account clustering, disinformation detection, and deepfake detection.
3. **Maintain and publish statistical norms against which inauthentic or coordinated behavior will be judged.** It is a hard problem to know what organic behavior on social-media topics and networks look like.
4. **Actively cultivate global participation in and feedback on this platform and these tools.** Subsidize participate in training and events from organizations and individuals from underrepresented regions. Solicit feedback on the platform and tools from those people, and compensate them for that feedback.

SUCCESS

Successful implementation of the above recommendations would result in standards and practice which will facilitate more robust, better integrated research and understanding. With access to a broader range of platforms, and an understanding of norms across platforms, these public-facing reports will present and more accurate picture of the real disinformation ecosystem. As standards of practice are adopted, comparing results across reports to draw synthetic conclusion will be easier, and tracking changes in bad behavior will be practical.

Current approaches are akin to trying to diagnose an illness without any patient history or prior case reports and by only looking at the head, heart, and lungs. Just as the development of the stethoscope or blood pressure cuff advanced diagnostic medicine, standard definitions, statistical metrics, and improved methodologies will help researchers identify and mitigate information operations on social media. In short, these changes will allow the study of disinformation to move from simply describing symptoms to actually meaningfully understanding the illness.

Figure 2. Share of Organizations with a Report Covering Each Platform

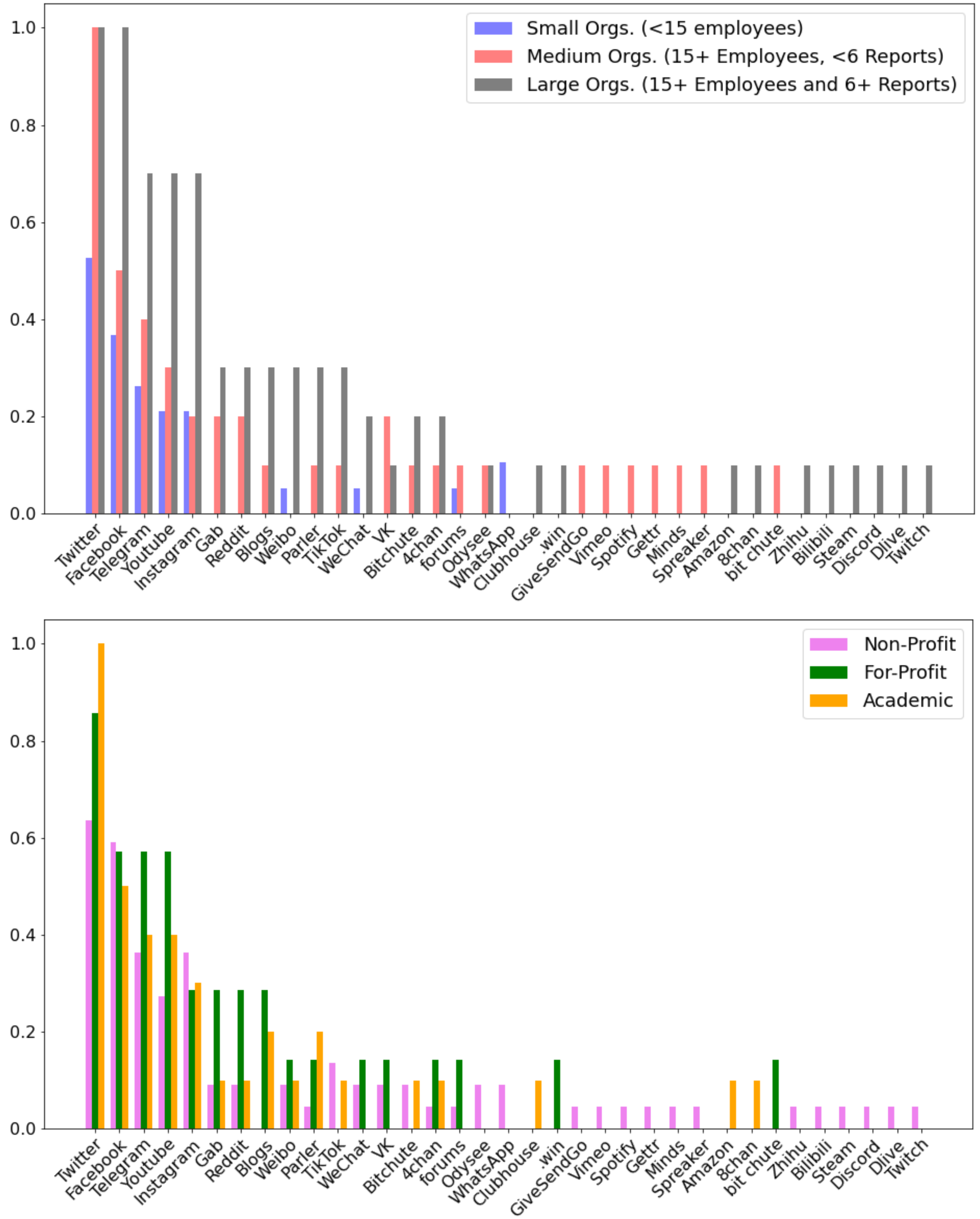


Figure 3. Share of Organizations Using Each Data Acquisition Method in a Report

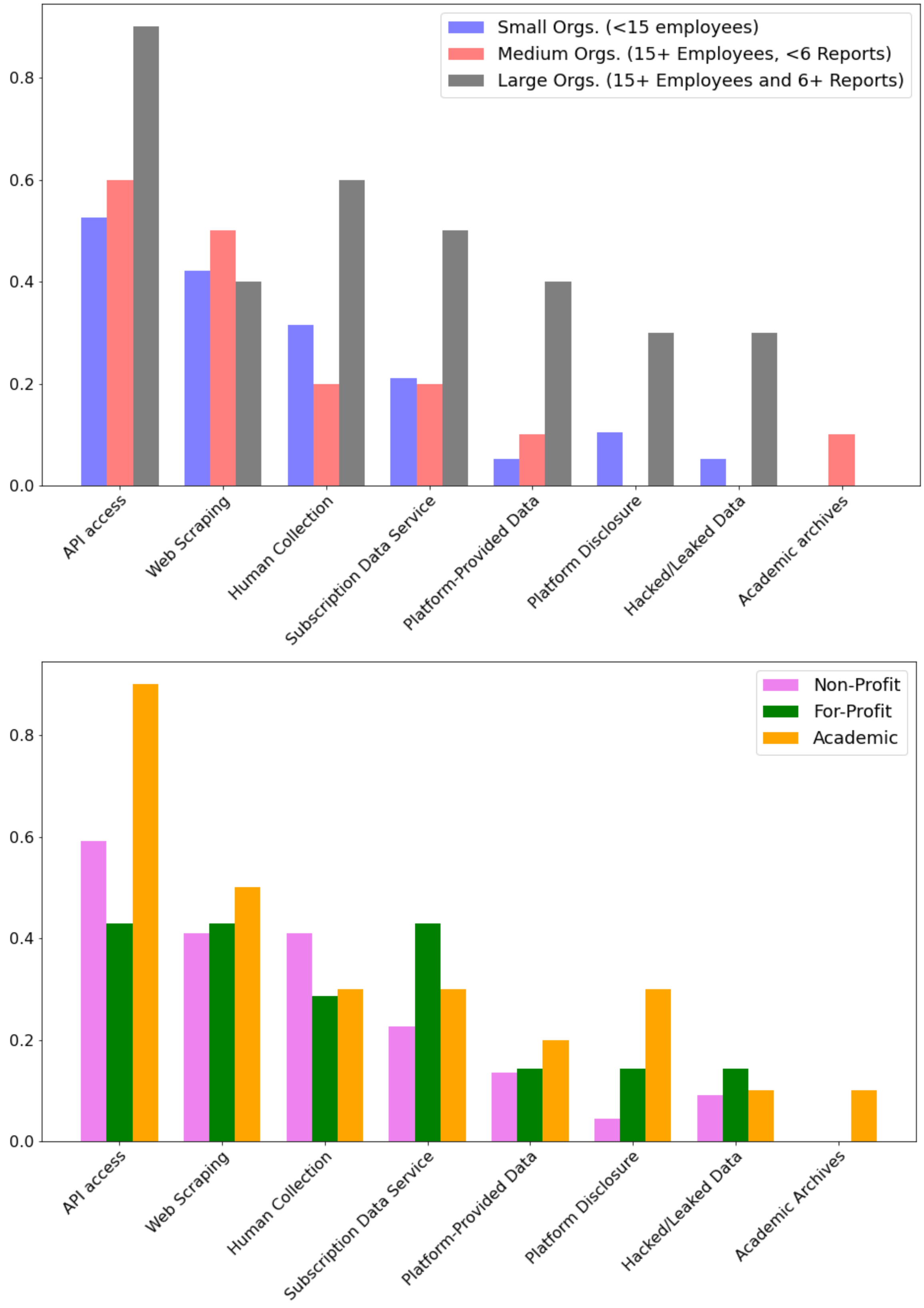


Figure 4. Share of Organizations Using Each Sort of Data in a Report

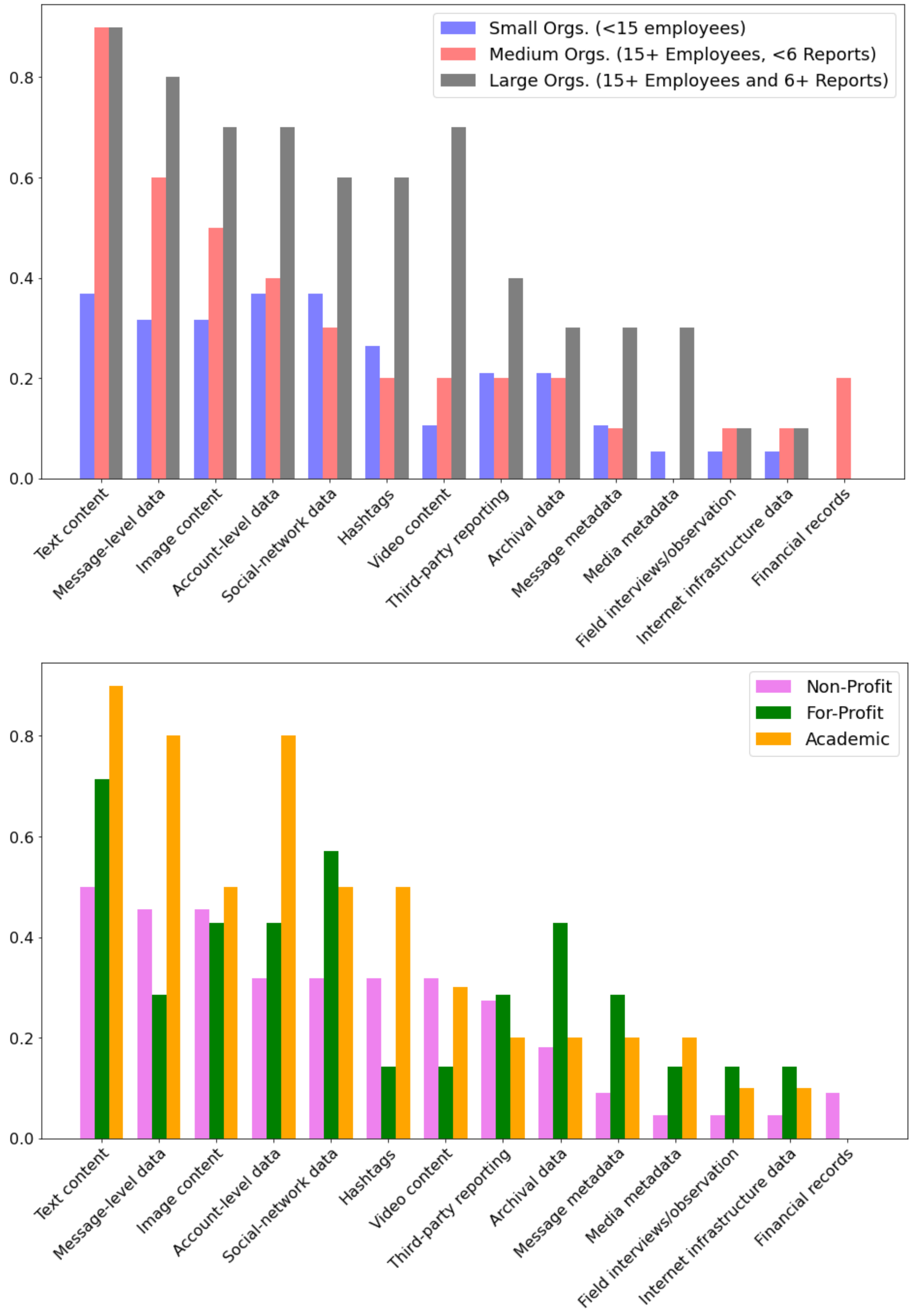


Figure 5. Share of Organizations Using Each Analysis Technique in a Report

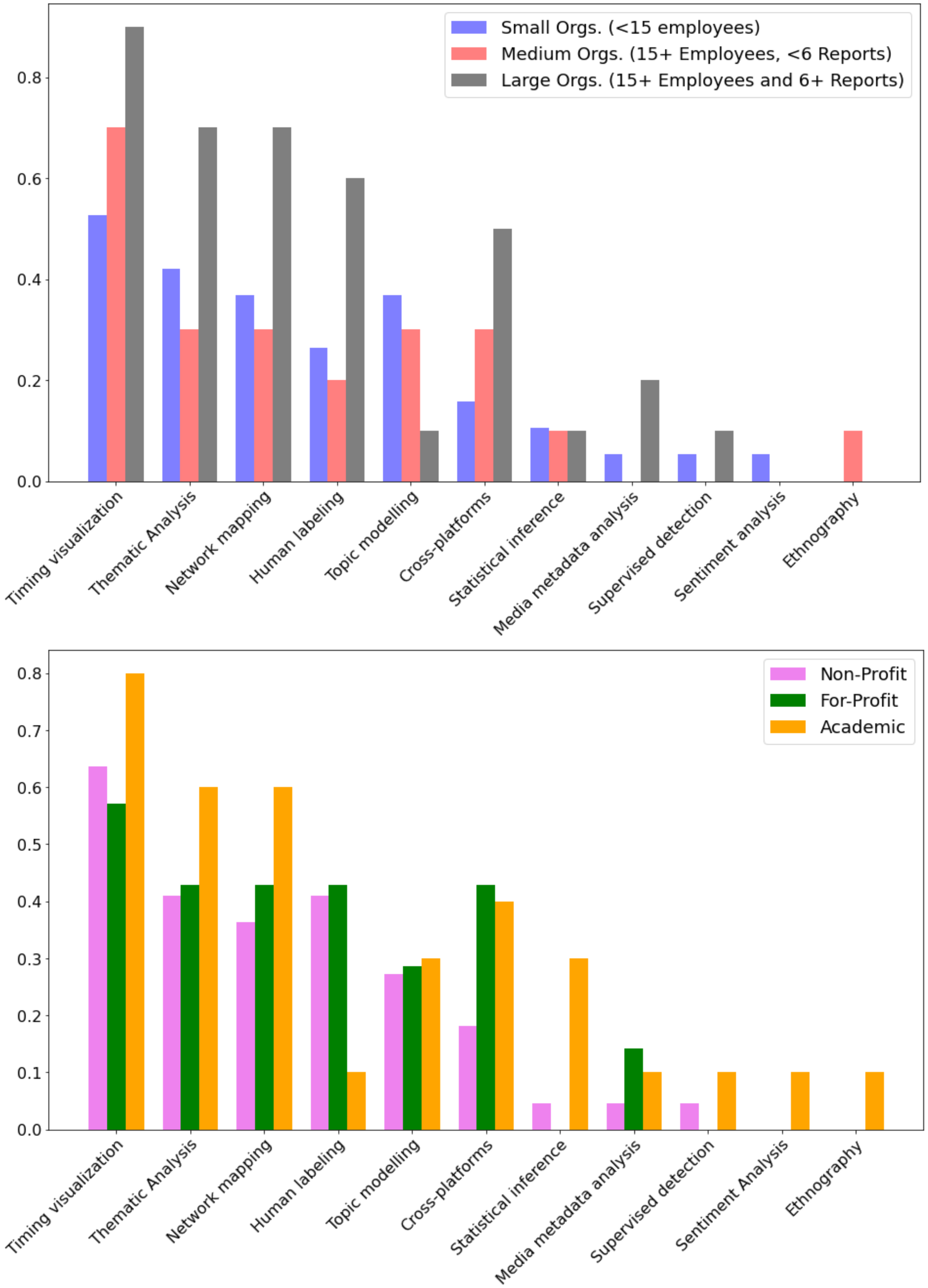


Figure 6. Share of Organizations Using Each Attribution Technique in a Report

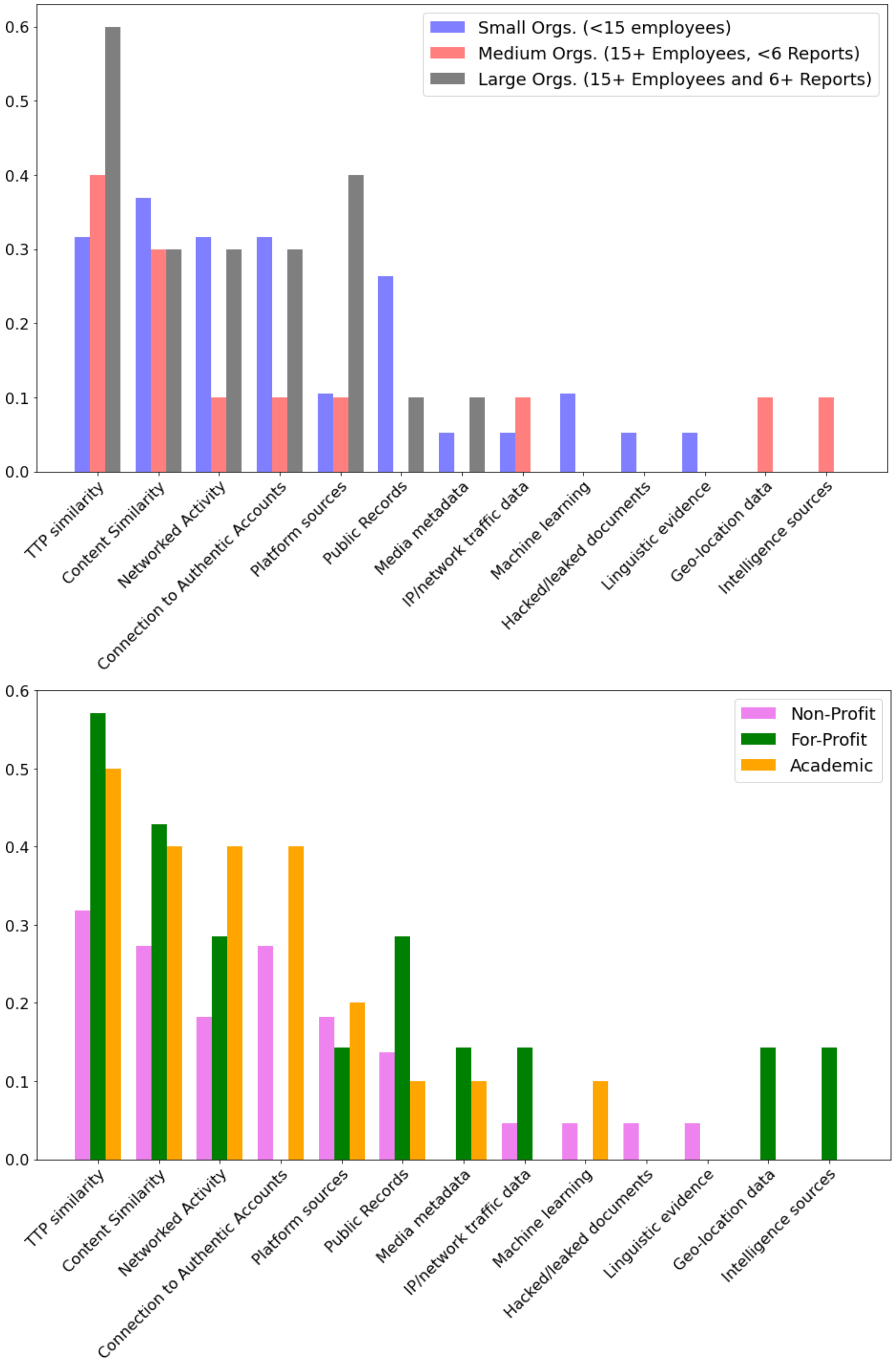


Figure 7. Share of Organizations with a Report Mentioning Each Limitation

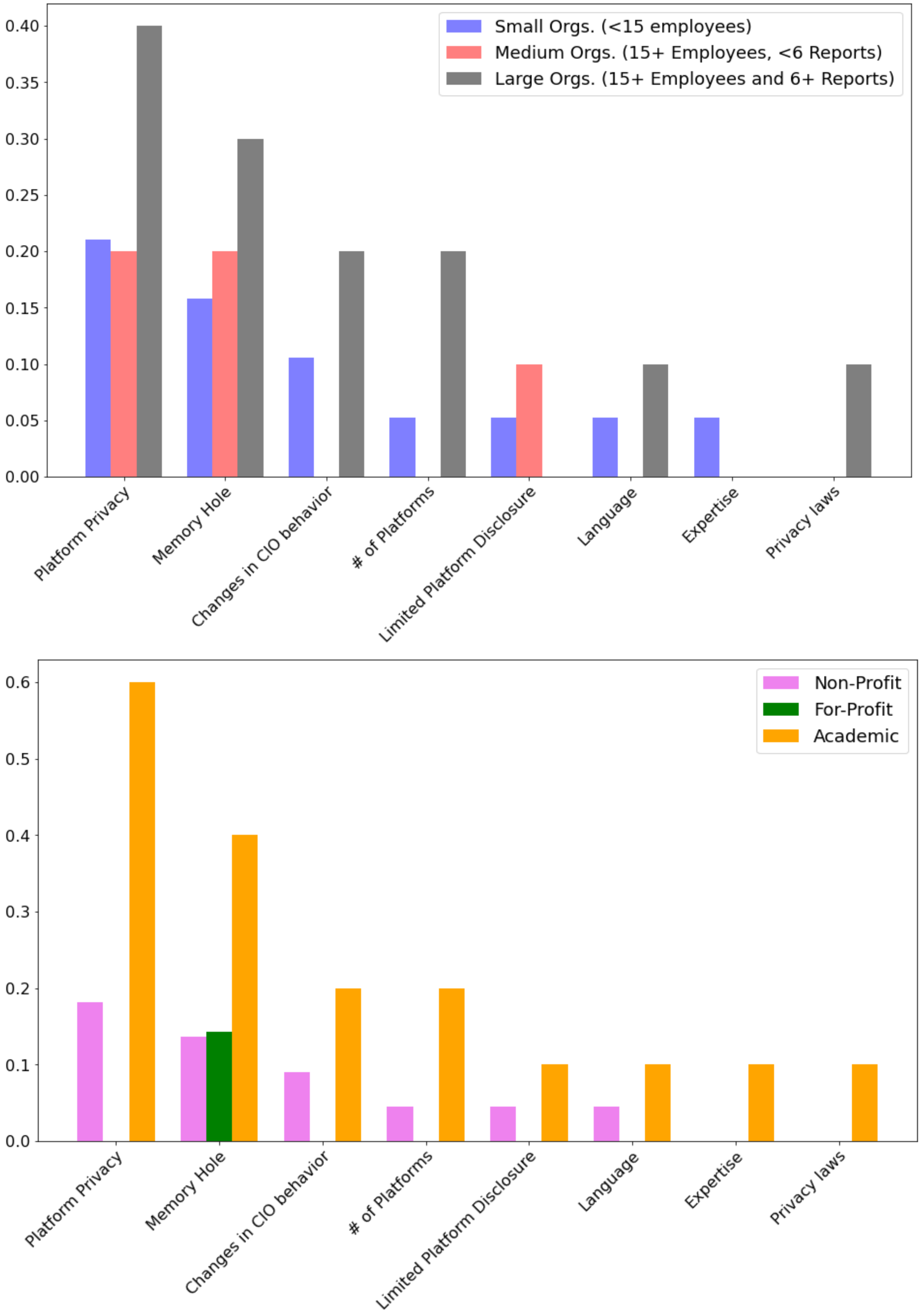


Figure 8. Share of Organizations with a Report Subject to Each Limitation

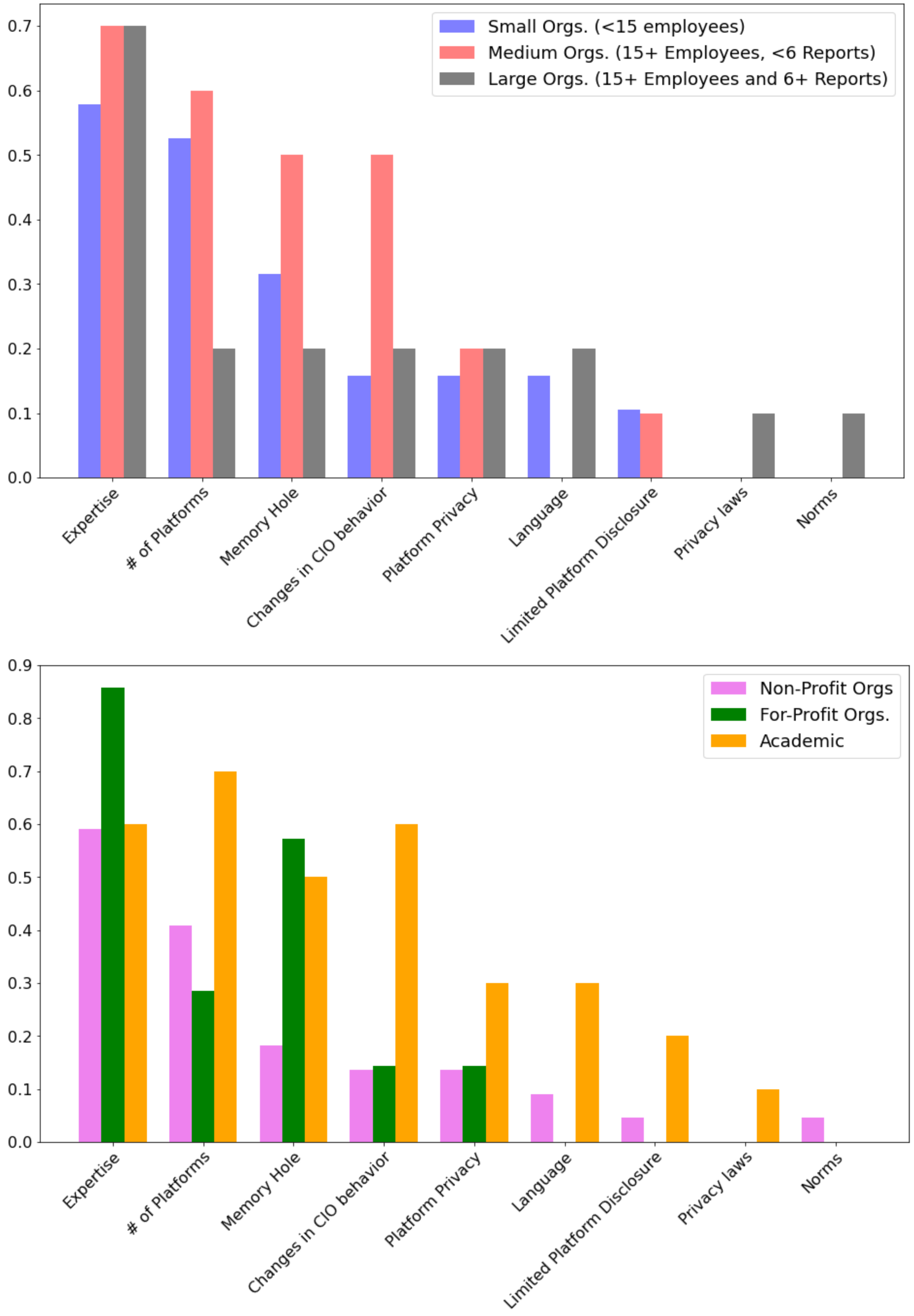
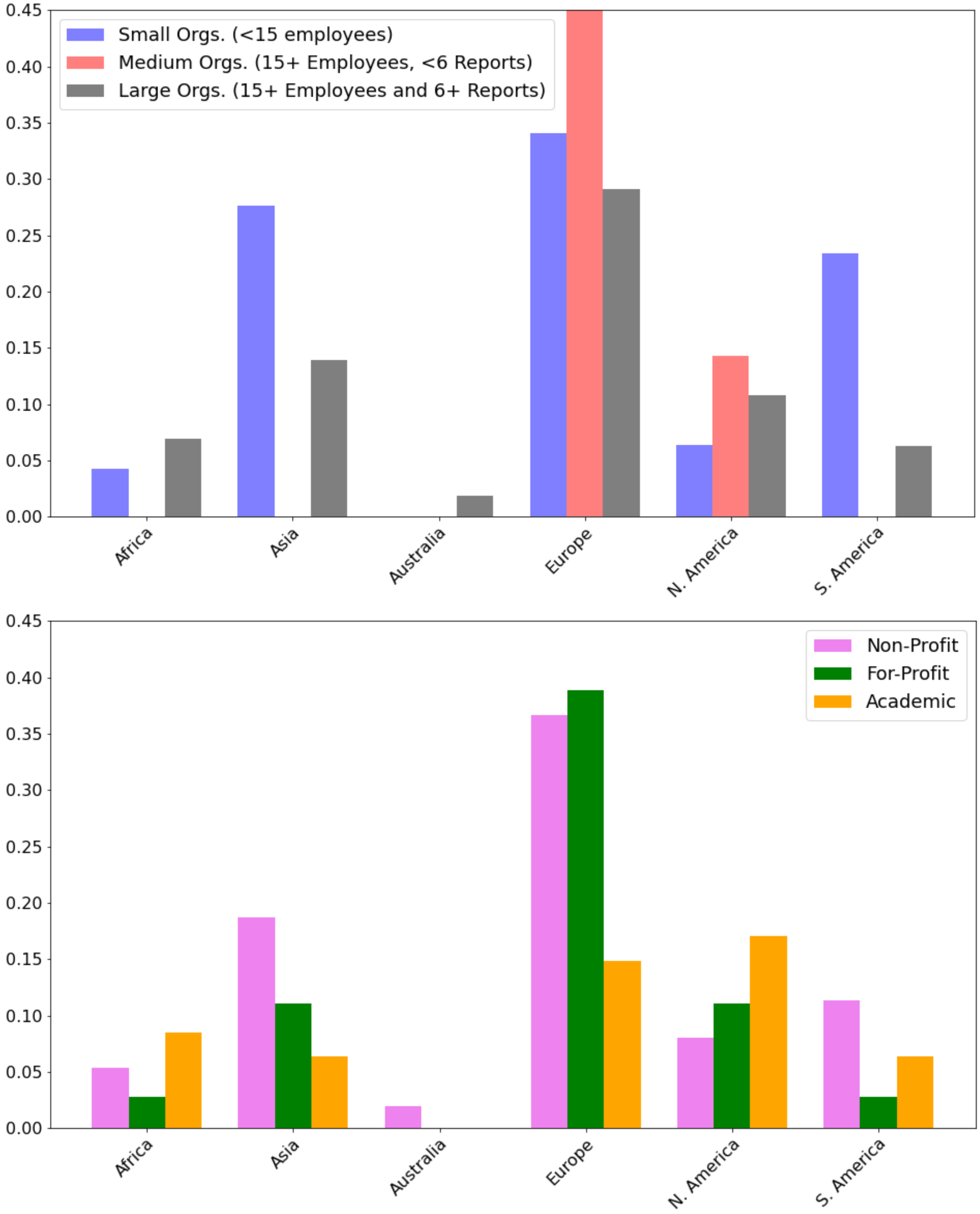


Figure 9. Share of Reports Covering Content Targeting Each Continent



References

Bradshaw, S., & Howard, P. N. (2019). The global disinformation order: 2019 global inventory of organized social media manipulation. *Computational Propaganda Research Project.* https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/09/CyberTroop-Report19.pdf

Linvill, D. L., & Warren, P. L. (2021). The real target of authoritarian disinformation. *Foreign Affairs.* *https://www.foreignaffairs.com/articles/russian-federation/2021-03-24/real-target-authoritarian-disinformation*

MacDonald, E. (2017, January 13). The fake news that sealed the fate of Antony and Cleopatra. *The Conversation.* https://theconversation.com/the-fake-news-that-sealed-the-fate-of-antony-and-cleopatra-71287

Nakov, P., Corney, D., Hasanain, M., Alam, F., Elsayed, T., Barrón-Cedeño, A., Papotti, P., Shaar, S., San Martino, G. (2021). Automated fact-checking for assisting human fact-checkers. *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI-21).*

Nunberg, G. (2019, December 30). ‘Disinformation’ is the word of the year—And a sign of what’s to come. *NPR*. https://www.npr.org/2019/12/30/790144099/disinformation-is-the-word-of-the-year-and-a-sign-of-what-s-to-come

Shane, S. & Mazzetti, M. (2018, September 18). The plot to subvert an election. *The New York Times.* https://www.nytimes.com/interactive/2018/09/20/us/politics/russia-interference-election-trump-clinton.html?mtrref=undefined&gwh=6E77CE0B25F3AFF73B1CA99AB3DB84D4&gwt=pay&assetType=PAYWALL

Silverman, C., & Kao, J. (2022, March 11). Infamous Russian troll farm appears to be source of anti-Ukraine propaganda. *ProPublica*. https://www.propublica.org/article/infamous-russian-troll-farm-appears-to-be-source-of-anti-ukraine-propaganda

Timberg, C. (2020, October 7). Parler and Gab, two conservative social media sites, keep alleged Russian disinformation up, despite report. *The Washington Post.* *https://www.washingtonpost.com/technology/2020/10/07/russian-trolls-graphika-parler-gab/*

Appendix A. Survey and Codebook

Data Collection for Princeton-Carnegie IO Research

1. Organization Name

____

2. Website

____

Organizational Detail

GO to the organizations "About Us" page or similar. According to that page:

3. Does This Org Analyze the Behavior of CIOs or other organized bad actors on social media for public consumption as a core element of their mission? Mark only one oval.

-Yes. Continue to next Page
-No. End Survey Now.

- **"Core element" means that it is not merely one of a large set of activities they pursue.**
- **The Actors need to be central to the analyzes, not just fact checking.**
- **"Public consumption" means reports that are accessible to non-experts and not meant for peer-review or technical audiences.**

4. Organizational Form

Check all that apply.

□ For-Profit

□ Non-Profit and Non-Academic

□ Academic

[Tag as academic if the org is hosted at a university. Tag as non-profit if org is registered as charity or similar in its home country. Tag as for-profit if the org has beneficial shareholders.]

5. Which of the following funding sources are acknowledged on the "About Us" section of the websites (or similar)

Check all that apply.

□ State Funding

□ Private Foundations

□ Clients

□ Platform

□ Corporate Donations Other:

□ ______________________________

6. Approximate Staffing

Mark only one oval.

○ less than 15

○ 15-50

○ 51+

[Include full-time and part-time staff, but do not include non-resident affiliates or collaborators. Use “About Us” page and Linked In profiles to estimate]

7. Copy/Paste Org Mission Statement

____

____

Report

To answer the following questions, review all reports that analyze the behavior of Coordinated Inauthentic Influence Operations, misinformation monetizers, violence coordinators, or other coordinated online disinformers, posted to or linked from the organization's website and to which this Org. contributed to since May 1, 2021. If there are no such reports, end the survey now.

[Find and download every qualifying report, define as a public-consumption report authored by the organization and linked to on the website or social media]

8. Number of reports reviewed:

9. In those reports, does this org make explicit claims of attribution?

Mark only one oval.

○ Yes

○ No

○ Maybe

[Claims like “We can never be completely confident, but… “ count as attribution. Ambiguous phrases like “Russian-aligned” count as a “Maybe”.]

10. What methods do they make to support attribution claims?

Check all that apply.

- □ Geo-location data
- □ TTP similarity **[Broad similarity between this campaign and known campaigns]**
- □ Hacked/leaked documents
- □ Insider disclosure **[Insiders to the campaign/actor group]**
- □ Linguistic evidence
- □ Content Similarity **[Text/Image/Video content]**
- □ Machine learning
- □ Networked Activity **[Connections between accounts in campaign]**
- □ Media metadata
- □ IP/network traffic data
- □ Intelligence sources
- □ Platform sources
- □ Public Records
- □ Connection to Authentic Accounts
- □ Other: ____________________

11. Platforms Analyzed

Check all that apply.

□ Twitter

□ Facebook

□ Instagram

□ TikTok

□ Youtube

□ Parler

□ Gab

□ Reddit

□ Blogs

□ LinkedIn

□ VK

□ Wikipedia

□ Weibo

□ Pinterest

□ Snapchat

□ Telegram

□ WeChat

□ Kuaishou

□ Qzone

□ Quora

□ .win

□ Other: ____________________

[Check all platforms from which at least one message/account/campaign was analyzed, even if not central to the report.]

12. Which of the following data acquisition processes does this organization use?

Check all that apply.

□ API access

□ Web Scraping

□ Subscription Data Service

□ Human Collection

□ Hacked/Leaked Data

□ Platform Disclosure **[Publicly Released]**

□ Platform-Provided Data **[Privately Provided]**

□ Other: ____________________

13. Which of the following sorts of data are analyzed in this organization's reports?

Check all that apply.

□ Message-level data **[Data from individual messages]**

□ Text content

□ Image content

□ Video content

□ Message metadata **[Data not directly accessible in normal viewing]**

□ Account-level data **[Data about individual accounts]**

□ Media metadata **[Data not directly accessible in normal viewing]**

□ Social-network data **[Network connections, defined any way]**

□ Hashtags

□ Archival data [**Archive.org, Google Cache, or similar]**

□ Third-party reporting

□ Internet infrastructure data

□ Field interviews/observation

□ Other: ____________________

14. Which of these Tools/techniques were used to analyze data?

Check all that apply.

□ Network mapping **[Visualizing connections between entities]**

□ Timing visualization

□ Topic modelling **[Separating messages into coherent subset on similar topics in an automated way]**

□

□ Statistical inference **[Applied in any capacity]**

□ Supervised detection **[Trained ML models]**

□ Thematic analysis **[Noticing and documenting major themes]**

□ Media metadata analysis

□ Cross-platforms **{Any attempt to trace operation across platforms]**

□ Human labelling **[Formal or ad-hoc labelling procedure on accounts or messages that included human judgement]**

Other: ______________________________

15. What features were explicitly cited as limiting their work?

Check all that apply.

□ Memory Hole **[That messages disappear from view for whatever reason]**

□ Limited to single or small # of platforms.

□ Privacy settings/characteristics of platforms

□ Privacy laws

□ Computing time

□ Expertise **[Unable to apply certain Technical, Historical, or Social-Scientific methods because they require skills or expertise that the authors lack.]**

□

□ Language **[Difficulty in interpreting messages due to language]**

□ Platform security team opacity **[Dependent or choices, methods, or judgement of platform teams without know how those were decided.]**

□ Changes in CIO behavior

Other: ______________________________

16. What elements seem to be limiting this organizations ability to work in this area?

Check all that apply.

□ Memory Hole **[That messages disappear from view for whatever reason]**

□ Limited to single or small # of platforms.

□ Privacy settings/characteristics of platforms

□ Privacy laws

□ Computing time

□ Expertise **[Unable to apply appropriate Technical, Historical, or Social-Scientific methods**
□ **because they require skills or expertise that the authors lack or are unaware of.]**

□ Language

□ Platform security team opacity **[Dependent or choices, methods, or judgement of platform teams without know how those were decided.]**

□ Changes in CIO behavior Other:

[For each organization, mark each limitation if there is at least one report that suffers from each limitation and the authors do not note that limitation. To judge whether element is a limitation, ask whether the report would likely be substantially stronger if there were no constraints along that dimension.]

Appendix B. Included Organizations

Alliance for Securing Democracy	https://securingdemocracy.gmfus.org/
Australian Strategic Policy Institute	https://www.aspi.org.au/program/international-cyber-policy-centre
Cazadores de Fake News	https://www.cazadoresdefakenews.info/
CEDMO	http://cedmohub.eu
Center for Countering Digital Hate	https://www.counterhate.com/
ClemsonHub	https://www.clemson.edu/centers-institutes/watt/hub/index.html
CSMAP	http://csmapnyu.org
DECACTO	https://defacto-observatoire.fr/Main/#
DFRLab	https://www.digitalsherlocks.org/about
Disinfo Defense League	https://www.disinfodefenseleague.org/
Doublethink Lab	https://doublethinklab.org/
EDMO BELUX	https://belux.edmo.eu/
EDMO Ireland	https://edmohub.ie/
Election Integrity Partnership	https://www.eipartnership.net/
EU Disinfo Lab	https://www.disinfo.eu/
First Draft News	https://firstdraftnews.org/
Global Disinformation Index	http://www.disinformationindex.org
Graphika	https://www.graphika.com/
Iberifier	https://iberifier.eu/
IDMO	https://www.idmo.it/
Institute for Strategic Dialogue	http://www.isdglobal.org
Luiss Datalab	https://datalab.luiss.it/
Mandiant	https://www.mandiant.com/
Miburo	https://miburo.com/
Moonshot	http://www.moonshotteam.com
Myth Detector	https://mythdetector.ge/en/
NORDIS	https://datalab.au.dk/nordis
Open Source Communications, Analytics Research	http://upsi.org.uk/oscar/
Prague Security Studies Institute	https://www.pssi.cz/
Pro Box	https://proboxve.org/
Programme on Democracy & Technology	https://demtech.oii.ox.ac.uk
Shorenstein Center -- Technology and Social Change	https://shorensteincenter.org/programs/technology-social-change/
Signa Lab	https://signalab.mx/
Social Observatory for Disinformation and Social Media Analysis	https://www.disinfobservatory.org/
Stanford Internet Observatory	https://cyber.fsi.stanford.edu/io
Tattle	https://tattle.co.in/
UW - Center for an Informed Public	https://www.cip.uw.edu/

VerificadoMX	https://verificado.com.mx/
VSquare	Vsquare_Project