TRUSTLAB

Code of Practice on Disinformation

A Comparative Analysis of the Prevalence of Misinformation and Sources of Disinformation across Major Social Media Platforms in Poland, Slovakia, Spain, and France

The Code of Practice on Disinformation was established following the European Commission's guidance by major online platforms, emerging and specialised platforms, players in the advertising industry, fact-checkers, research, and civil society organisations to deliver a strengthened Code of Practice on Disinformation. This measurement study was commissioned by Meta, YouTube, TikTok, and LinkedIn as part of the European Commission's Code of Practice. You can read more about the Code of Practice framework here.

Table of Contents

Executive Summary

The European Union's (EU) 2022 Code of Practice on Disinformation represented a milestone in the global fight against online disinformation. Online disinformation is an ambiguous and fast-changing phenomenon, and measuring disinformation is challenging. As the second iteration of the Code, this study set out to evaluate the prevalence and sources of disinformation across five major social media platforms (Facebook, Instagram, LinkedIn, TikTok, and YouTube) in four countries: Poland, Slovakia, Spain, and France. A total of 18,385 social media posts and 10,654 accounts were sampled by searching popular disinformation keywords using a platforms' native search functionality during the period May to June 2024, a key time for European elections. The key metrics examined are discoverability, relative post engagement, absolute post engagement, and properties about disinformation actors, including ratio of disinformation actors, their account activities, and engagement with other users.

- Discoverability refers to the percentage of content returned from searching disinformation keywords. It captures how easily a platform surfaces misinformation or disinformation content to a user searching for sensitive topics. The platform with the largest discoverability was Facebook. YouTube had the lowest ratio of discoverability.
- The most common misinformation category we found concerned General Misinformation; this was followed by Politics and Current Events defined as misinformation concerning political events, governmental actions, or current affairs. Manipulated or Synthetic Media was the least common misinformation category we observed, defined as content involving media that has been altered, synthesised, or manipulated to deceive or misinform.
- Overall, Spain was found to have the highest discoverability ratio of the countries we studied. Slovakia had the lowest discoverability rate.
- Overall, we examined 10,654 social media accounts to determine if they were disinformation actors, and we identified 652 disinformation actors (6%). The number of disinformation actors identified as a fraction of all accounts investigated was found to be highest on Facebook (12%), followed by Instagram (9%) and TikTok (5%). YouTube (2%) and LinkedIn (2%) had the lowest fraction of disinformation actors.
- Posts written by disinformation actors received less engagement compared to posts written by non-disinformation actors on all platforms. Facebook had the highest relative engagement ratio and YouTube had the lowest. This same relative ordering was observed across all three measurements we conducted, except for one instance where LinkedIn narrowly surpassed TikTok in Measurement 1.
- Platform-supplied labels are exceedingly rare, such that 95% of the posts we ascertained to contain misinformation were not actioned by the platform.

Several improvements were made to the methodology following the pilot study in 2023. For example, the sample size was increased by 3x, engagement metrics were improved, platform visible actions were recorded, and subcategory labels were added.

About the Author

TrustLab's mission is to build a safer internet for everyone.

It is a global leader in disinformation monitoring and analytics, and serves as a trusted third-party solution for detecting and mitigating critical safety threats on the internet. Its three founders have over forty years of combined Trust & Safety experience from companies including Google, YouTube, Reddit, and TikTok.

TrustLab has worked with many social media platforms, messaging companies, government bodies, and online marketplaces to deliver its innovative, independent, and unbiased measurement solution. Leveraging state-of-the-art, patent-protected technology, it is able to accurately and rapidly identify harmful content at scale across multiple languages, sources and abuse verticals.

TrustLab's customers range from small companies building out their internal teams and policies to large enterprises with complex Trust & Safety needs. By providing its cutting-edge software and expertise, TrustLab helps its clients protect their users against harmful content and in doing so, serve its mission to make the internet safer.

About the Participants and Partners

The participants in the study were five major social media platforms (Facebook, Instagram, LinkedIn, TikTok, and YouTube).

Facebook is an online social media and social networking service owned by Meta Platforms.

Instagram is a photo and video-sharing social networking service owned by Meta Platforms.

LinkedIn is a business and employment-focused social media platform owned by Microsoft.

TikTok is a short-form video entertainment platform owned by ByteDance.

YouTube is an online video-sharing and social media platform owned by Google.

The partners of the pilot study were the Permanent Taskforce of the Code of Practice, in particular the European Commission, Avaaz, European Regulators Group for Audiovisual Media Services (ERGA), and European Digital Media Observatory (EDMO).

The **European Commission** helps to shape the European Unions' overall strategy, propose new EU laws and policies, monitor their implementation and manage the EU budget. It also plays a significant role in supporting international development and delivering aid.

Avaaz is a global web movement to bring people-powered politics to decision-making everywhere.

The **European Regulators Group for Audiovisual Media Services (ERGA)** brings together heads or high level representatives of national independent regulatory bodies in the field of audiovisual services, to advise the Commission on the implementation of the EU's Audiovisual Media Services Directive (AVMSD).

The **European Digital Media Observatory (EDMO)** brings together fact checkers, media literacy experts, and academic researchers to understand and analyse disinformation, in collaboration with media organisations, online platforms and media literacy practitioners.

About the Structural Indicators

Under Commitment 41 of the Code, signatories commit to working within the Task-force to develop Structural Indicators designed to assess the effectiveness of the Code in reducing the spread of online disinformation for each relevant signatory and for the entire online ecosystem in the EU and at the Member State level.

To achieve this, signatories established a Working Group in June 2022 following the launch of the Strengthened Code and the European Commission requested EDMO to create a first proposal for Structural Indicators to initiate discussions within the Working Group. EDMO presented a proposal at the beginning of September 2022, encompassing six different areas: prevalence, sources, audience, demonetisation of disinformation, collaboration, and investments in fact-checking and Code implementation. Due to the comprehensiveness of EDMO's proposal and the limited time available, the Working Group and EDMO agreed to focus on the prevalence, sources and audience of disinformation as the initial set of Indicators for the 2024 reporting.

While in the course of autumn of 2022, several platform signatories had worked towards significantly increasing their data point availability on the prevalence and sources of disinformation, the tabled datasets and data points did not allow for satisfactory cross-platform Structural Indicators. Platform signatories noted that they had done their utmost to meet the Working Group's timelines and accommodate said data requests, taking into account (legal) constraints.

In January 2023, platforms committed to evaluating whether one or more third parties should be selected to assist in delivering the first set of Structural Indicators, either independently or with the support of EDMO, by the first reporting period. To ensure a harmonised approach across the main platforms (Facebook, Instagram, LinkedIn, TikTok, and YouTube) and to adhere to their Terms of Service, the Working Group issued a call for proposals and decided to contract TrustLab for the pilot independent analysis of the selected indicators. The pilot was completed and published in September 2023, focused on Spain, Poland, and Slovakia.

In 2024, having explored ways to expand the scope and methodology, for this second study, a fourth EU Member State, France, has been covered. Furthermore, a number of methodological improvements have been made, which are detailed in this report.

Terminology

The following definitions are based on an amalgamation of peer-reviewed studies that TrustLab considers to be broadly aligned with industry standards that were then adopted for TrustLab's policies. These definitions are aligned with the European Union's 2022 Code of Practice on

Disinformation; however, "foreign interference in the information space" is outside the scope of the current study. Better access to platform data and alignment on operational definitions can enable future measurements to address this limitation.

Term	Definition
Misinformation	Misinformation is false or misleading content shared without the intent to deceive.
Disinformation	False or misleading information that is intentionally created and spread to deceive or manipulate others. This must be a deliberate dissemination of false narratives, fabricated stories, or manipulated facts with the intention to cause harm, manipulate public opinion, or achieve a specific agenda. This is a simplification of the EC definition ("false or misleading content that is spread with an intention to deceive or secure economic or political gain, and which may cause public harm") and reflects what was operationalised in the study. The intent of the actor, nature of gain they hope to receive, and how much public harm can be (or is intended to be) caused are not readily visible from the content itself, and need to be inferred, with ample room for subjectivity. To operationalise the measurement of disinformation, we focused on visible signs from the user who posted the content such as (but not limited to) repeat activity, size of the follower network, manipulation of images, video, or audio clips, the deliberate use of misleading headlines, or clickbait as a way to attract attention and promote false narratives.
Disinformation Actors/ Disinformation Campaign	Accounts actively posting disinformation. This determination is reached once the account has been reviewed to acquire further information. The Disinformation Actor "ABCDE" framework (initially suggested by the Carnegie Endowment for International Peace) is applied to determine the nature of the account, analyse the users' behaviour, content determination, evaluate the primary audience, and effect of the content being shared. Once content is determined to be misinformation, the account responsible for posting it is sent for disinformation actor review.
Irrelevant	Content that is completely irrelevant to the Code of Practice, devoid of any misinformation narrative. Content that may be perpetuating what most people would consider "untruthful" but framed in the context where no reasonable person would believe it. Additionally this includes content that serves educational, documentary, scientific, or artistic purposes, emphasising factual and non-misleading information.

Term	Definition *(continued)*
Misrepresentation	Content that dramatically portrays misinformation within an entertainment context, such as Op-Ed coverage, where it may be misconstrued as unbiased news or factual misinformation.
Potentially Harmful	Content that glamorises or gratuitously depicts misinformation, posing a significant potential for harm to the public.
Harmful and Dangerous	Content that is demonstrably untrue or is intentionally deceptive with a direct link to user or societal harm.

The following definitions relate to the metrics used by TrustLab in this study.

Metric	Definition
Discoverability	The percentage of content returned from searching disinformation keywords. It captures how easily a platform surfaces misinformation or disinformation content to a user searching for sensitive topics.
Engagement	Reactions, comments, shares, and views, depending on platform availability.[1]
Relative Post Engagement	The ratio of misinformation or disinformation engagement (where the underlying content is misinformation or disinformation) to non-misinformation or disinformation engagement (where the underlying content is non-misinformation or disinformation content).
Absolute Post Engagement	The magnitude, in absolute terms, of engagement with misinformation or disinformation content (with the caveat that the underlying data availability and nature across platforms can affect the magnitude of the metric).
Ratio of Disinformation Actors By Platform	The proportion of disinformation actors relative to the total accounts sampled on a platform.
Engagement With Disinformation Actors	The ratio of the engagement of disinformation actors with other users over the engagement of non-disinformation actors with other users on the platform. Absolute comparison is also provided. This sheds light on the influence that disinformation actors may exert on other users.
Disinformation Actor Account Activities	The group differences between disinformation and non-disinformation actors in post frequency and network size.

[1] Number of views is the engagement statistic that YouTube displays on the search results view. Theoretically it is possible to collect the number of reactions and the number of comments on YouTube as well; however, this information is only displayed on the video page itself. During data collection we do not visit this page and so do not have access to this data.

The following terms relate to the sub-categories used by TrustLab in this study.

Term	Definition
Civic and Election Integrity	Misinformation narratives related to the ethical adherence to principles, transparency, and fairness in civic processes, and ensuring the accuracy and transparency of elections.
Medical Misinformation	Narratives that contain inaccuracies or false claims about medical treatments, remedies, vaccines, diseases, and medical advice, which can lead to adverse effects on public health.
Manipulated or Synthetic Media	Content involving media that has been altered, synthesised, or manipulated to deceive or misinform.
General Misinformation	Covers a wide range of misinformation not falling under specific categories but is still demonstrably false.
War / World Conflicts	Misinformation narratives related to armed conflicts globally or domestically, alongside significant geopolitical events that escalate tensions between nations or within regions.
Hate-Based Misinformation	Misinformation that fosters prejudice, discrimination, or hostility towards individuals or groups based on their belonging to Protected Groups.
Politics and Current Events	Misinformation related to governmental affairs, political events, and significant global or domestic occurrences, excluding content specifically related to elections or civic processes.
Sensitive Topics	False or deceptive narratives surrounding delicate or controversial subjects, with the intent of exploiting emotional responses.

To determine if an account is a disinformation actor, TrustLab used the following formula:

Criteria[2]	Determination
If the account meets none or 1 of the criteria	The account is irrelevant to the project
If the account meets a combination of 2 criteria	The account is unlikely to be a disinformation actor.
If the account meets a combination of 3 criteria	The account has a low probability of being a disinformation actor.
If the account meets a combination of 4 criteria	The account is likely a disinformation actor.
If the account meets a combination of 5+ criteria	The account is a highly probable disinformation actor.
If the account is confirmed by a reputable third-party source to be actively involved in disinformation activities	The account is a verified disinformation actor.

[2] When reviewing accounts to determine whether or not the account is a disinformation actor, certain criteria must be evaluated. See more details in the ‘Improvements following Pilot Study’ section.

Methodology

The structural indicators proposed by EDMO ideally require collecting random samples of all content and accounts on platforms. This study, however, took a different sampling approach due to limitations around data collection such as a lack of access to internal platform data. The data underlying this study were collected automatically using third party data brokers. These enabled us to capture data samples in a timely and cost-effective manner, while avoiding human error and transcription mistakes.

Specifically, the sampling approach in this study was to search for specific keywords related to current misinformation topics on each platform's native search engine, and collect the posts and accounts from the search results. The collected data does not represent a random sample of all content and accounts on that platform, but instead represents the content and accounts encountered by users of the platform who are searching for keywords related to misinformation topics.

To understand a key difference between the two sampling approaches, consider that when normalising the amount of vaccine misinformation, the denominator can either be all content on the platform or only vaccine related content. Both measures are meaningful, and offer different perspectives on the prevalence of misinformation.

Step 1: Keywords were generated using fact check / trusted sources

Step 2: Data collection was carried out in three measurements over a period of six weeks[3]

Step 3: Data labelling was conducted to identify misinformation content and disinformation actors

The same number of posts were labelled for each country. We produced an initial data set of posts by searching for all keywords on all platforms; this was then used to produce balanced datasets for each country by constructing sets of search results that were "paired" across all four languages. This ensures that the posts selected for each country are distributionally similar in terms of search result rank and number of keywords represented. While this balancing means that the Spanish sample is less representative of Spanish content overall than the Slovakian sample is, an unbalanced procedure would have resulted in greatly reduced confidence in our estimates of Slovakian population metrics.

To ensure data collection quality, a multi-tiered review process was implemented. Tier-1 analysts are frontline content moderators responsible for handling the initial review of a piece of content. Tier-2 analysts are more experienced and specialised content moderators who deal with complex or nuanced issues that require expertise beyond the scope of tier-1 analysts. After tier-1 analysts reviewed the content, tier-2 analysts re-reviewed 20% of positive and negative content reviewed by the tier-1 analysts as a quality assurance. All analysts working on misinformation and disinformation annotation underwent comprehensive pre-project training. This included asynchronous eLearning, live sessions, blind reviews of confirmed cases, and ongoing calibration with subject matter experts to ensure consistent and

[3] Measurement 1 (13 May 2024 - 27 May 2024), Measurement 2 (27 May 2024 - 10 June 2024), Measurement 3 (10 June 2024 - 24 June 2024)

rigorous determinations. Further quality assurance measures included continuous agent performance monitoring, feedback sessions, and daily updates of the precision metric based on the tier-2 reviews.

Improvements following Pilot Study

The sample size was increased by 3x for social media posts and over 2x for social media accounts, which helped in attaining better statistical significance. The bigger sample size has facilitated better characterisation of sources of disinformation than what was possible in the pilot study.

Engagement metrics have been improved so that in addition to the combined engagement metric (reactions + comments + shares), we have split out those component metrics (such as reactions) that are available on all platforms to allow additional detailed analysis.

When measuring discoverability and engagement, we explored the resulting actions the platforms took against misinformation posts. This included warning labels[4], fact-check links, or other platform visible actions were recorded, so that discoverability and engagement could be split by misinformation content labelled as such, versus unlabeled content.

We applied more granular labels to allow for the splitting out of metrics such as discoverability, engagement, and sources of disinformation based on these sub-categories. We report statistics on the number of misinformation posts that we found that had been actioned by the platform (e.g., by including a warning label or fact check link). These platform-supplied labels validate our misinformation judgements, in that our verdicts of misinformation content accord with every instance where a post has a platform-supplied warning label or fact-check label. Conversely, platform-supplied labels are exceedingly rare, such that 95% of the posts we ascertained to contain misinformation were not actioned by the platform.

For disinformation actors, the intent and origin of the poster has been assessed using the "ABCDE" framework initially suggested by the Carnegie Endowment for International Peace[5]. Based on the framework, accounts sharing disinformation content have been categorised as one of: Unlikely Disinformation Actor, Low Probability Disinformation Actor, Likely Disinformation Actor, Highly Probable Disinformation Actor and Verified Disinformation Actor.

When reviewing accounts to determine whether or not the account is a disinformation actor, certain criteria[6] must be evaluated, including:

- Account creation time
- If there is a biography and/or profile picture
- If the account is repeating posts and comments (4+ times per day)
- If the account has under 100 friends or followers
- If the account is sharing and spreading altered, deep fake images or graphics
- If the account is not providing any personal updates or sharing personal information
- If the account is spreading conspiracy theories and hoaxes

[4] Warning labels and fact-check links were collected together with the posts at the same point in time. We did not conduct an additional review to ascertain whether warning labels or fact-check links were added to any posts after our initial collection.
[5] https://www.jstor.org/stable/pdf/resrep26180.6.pdf
[6] Please see the Terminology section for the formula

- If the account has assumed a persona with strong political views
- If the account is frequently sharing posts on the same topics (4+ times per day and 20+ times per month)

While our methodology provides valuable insights, it is not without limitations. The keyword-based approach may miss disinformation not captured by our selected keywords. Additionally, platform-specific search algorithms could influence the discoverability of content, introducing variability in our results. Future studies should aim for direct access to platform data for more comprehensive analysis, including of recommendation algorithms.

Key Findings

Please see the dashboard for in depth findings with supporting visualisations.

During the three measurements, we collected 18,385 social media posts (~4,600 posts from each of the four countries we investigated) and labelled whether these contained misinformation and disinformation narratives or not. Overall, we identified 3,985 misinformation posts, a ratio of 22%[7]. Since these posts were found by intentionally searching for keywords likely to surface misinformation, we call this ratio the "discoverability".

We measured misinformation and disinformation content to be most discoverable on Facebook (discoverability of 44%), followed by Instagram, TikTok, and LinkedIn; Youtube had the lowest discoverability at 9%. This relative ordering was consistent across all three measurements, indicating that our observations are repeatable and our analysis is robust.

Overall, Spain was found to have the highest discoverability ratio of the countries we studied. Slovakia had the lowest discoverability rate. This ordering was also consistent across all three measurements.

The most common misinformation category we found concerned General Misinformation (879 posts); this was followed by Politics and Current Events (694), Medical Misinformation (692), War / World Conflicts (651), Sensitive Topics (527), Hate-based Misinformation (284), and Civic and Election Integrity (197). Manipulated and Synthetic Media was the least common misinformation category we observed (61).

Overall, we examined 10,654 social media accounts to determine if they were disinformation actors, and we identified 652 disinformation actors (6%). The number of disinformation actors identified as a fraction of all accounts investigated was also found to be highest on Facebook (12%), followed by Instagram (9%) and TikTok (5%). YouTube (2%) and LinkedIn (2%) had the lowest fraction of disinformation actors.

Relative engagement with posts authored by disinformation actors was less than one on all the platforms we studied; that is, posts written by disinformation actors received less engagement compared to posts written by non-disinformation actors on all platforms. Facebook had the highest relative engagement ratio and YouTube had the lowest. This same relative ordering was observed across all three measurements we conducted, except for one instance where LinkedIn narrowly surpassed TikTok in Measurement 1.

Platform warning labels were found on 2% of posts (379 TikTok posts, and 22 Instagram posts). 9% of warning labels were applied to videos depicting physically dangerous activities (stunts, pranks) likely to cause bodily harm. Of the remaining warning labels, we observed very good agreement with TrustLab's determinations about posts containing misinformation:

---

[7] 95% confidence interval is (3876, 4095). The 95% confidence intervals were calculated for the discoverability and disinformation actor ratio metrics; these intervals were computed for the population proportions for those two metrics using the assumption of normality.

Interval := SampleMean +/- z * Sqrt((SampleMean) * (1 - SampleMean) / (SampleSize))

- *Sensitive content* (9 posts were determined by TrustLab to contain misinformation, representing 4% of the content found with this warning label)
- *Learn more about COVID-19 vaccines* (22 misinformation posts, 20%)
- *Learn about making informed financial decisions* (8 misinformation posts, 26%)
- *Missing Context. The same information was reviewed by independent fact-checkers in another post*. (11 misinformation posts, 100%)
- *False Information: The same information was reviewed by independent fact-checkers in another post*. (5 misinformation posts, 100%)
- *False information: Reviewed by independent fact-checkers* (3 misinformation posts, 100%)
- *Get info on the UK General Election* (2 misinformation posts, 66%)

As aforementioned, platform-supplied labels are exceedingly rare, such that 95% of the posts we ascertained to contain misinformation were not actioned by the platform.

In all, 50 TikTok posts containing misinformation and 19 Instagram posts containing misinformation had warning labels attached by their respective platforms. TrustLab found 813 misinformation posts on TikTok and 448 misinformation posts on Instagram that did not have warning labels attached. That is, 94% of misinformation posts found on TikTok and 96% of misinformation posts found on Instagram did not have warning labels.

Fact-check links were found on a further 15 Facebook posts. All of these posts were judged by TrustLab to contain misinformation narratives. TrustLab found another 1,902 posts on Facebook that were judged to contain misinformation, and which did not carry a fact-check link; that is, 99% of the misinformation posts found by TrustLab did not have fact-check links.

Conflicts and Acknowledgements

Although all measured platforms funded the study, TrustLab independently developed measurement methodology and conducted post measurement analysis, without platform input.

TrustLab would like to thank the Permanent Task-force of the Code of Practice, in particular the European Commission, Avaaz, European Regulators Group for Audiovisual Media Services (ERGA), European Digital Media Observatory (EDMO), and the platform signatories.

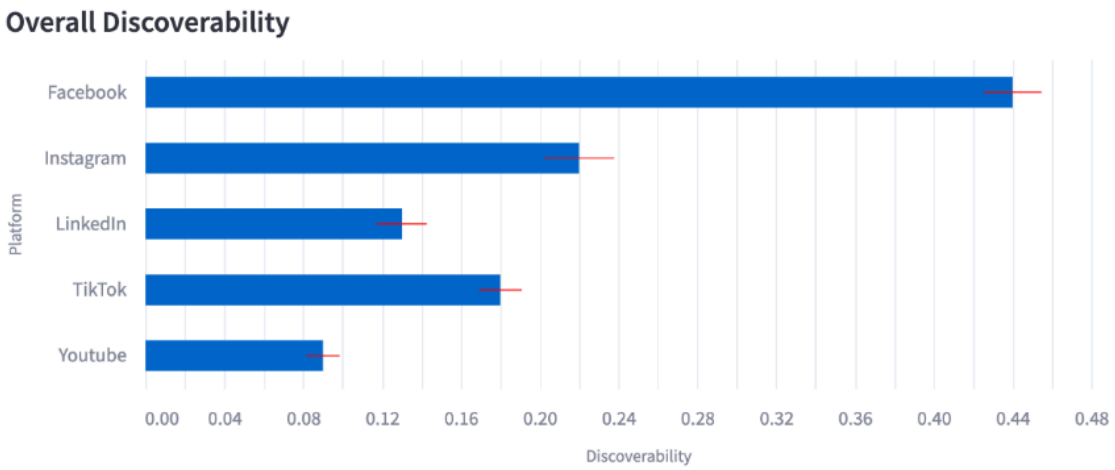
Appendix

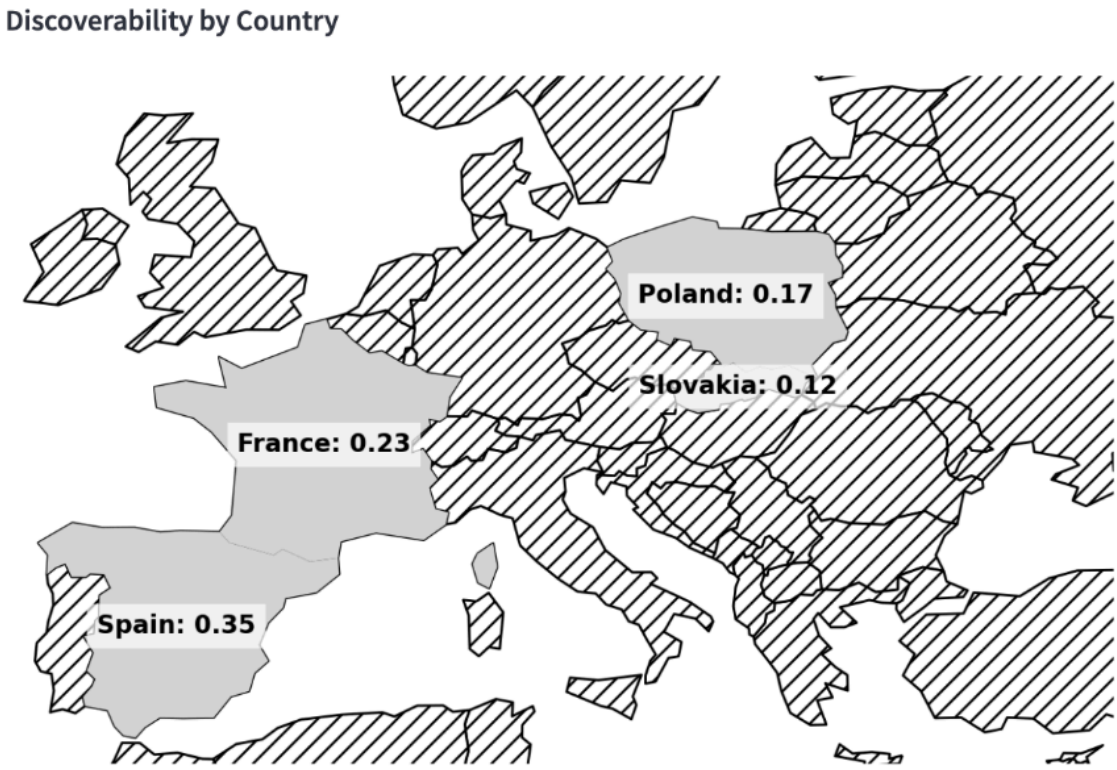
Example Keywords

Market	Example keywords
Poland	niemieckie PO, dyktatura UE, szczepienia autyzm
Slovakia	Green deal, #mimovladky, protokoly
Spain	Ojo Con las vacunas para tu mascota, proyecto blue beam, repentinitis
France	laboratoire biologique, Macron annonce sa démission, etatsunis effondrement

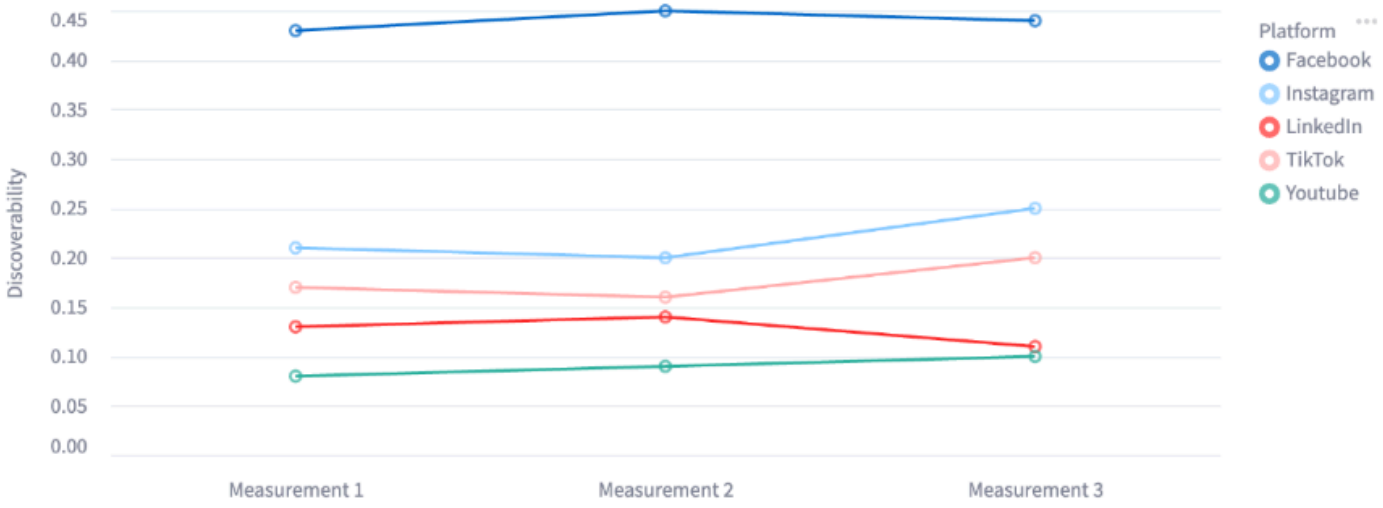
Graphs & Data Visualizations

1. Discoverability

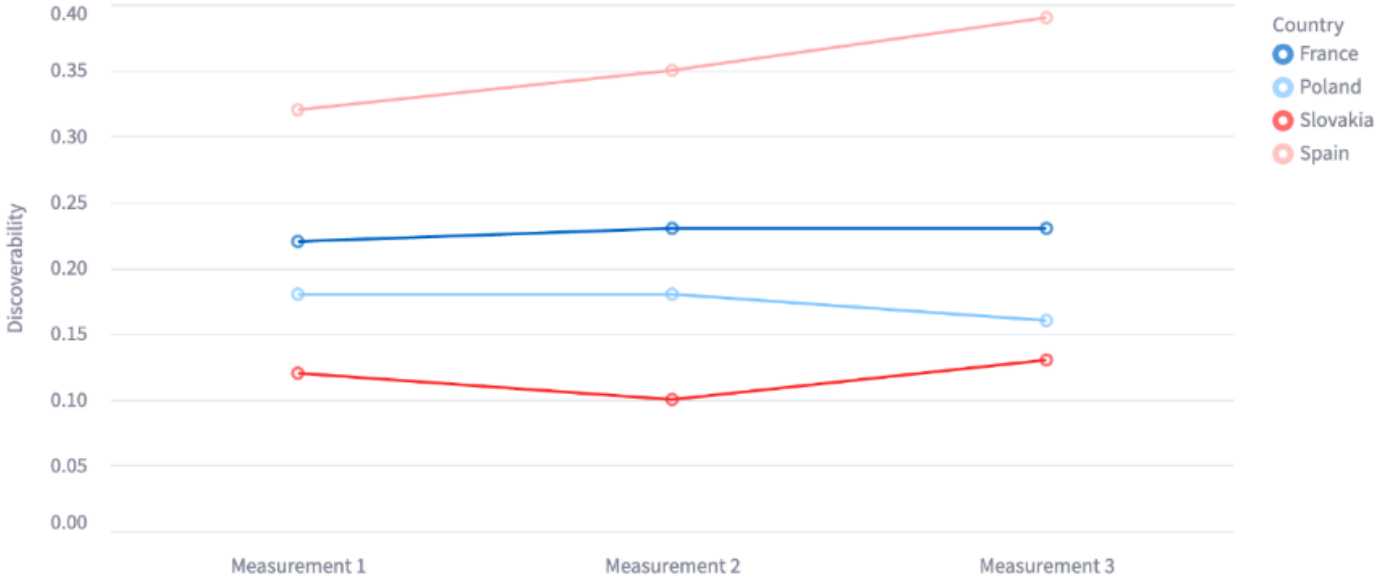




Discoverability by Measurement



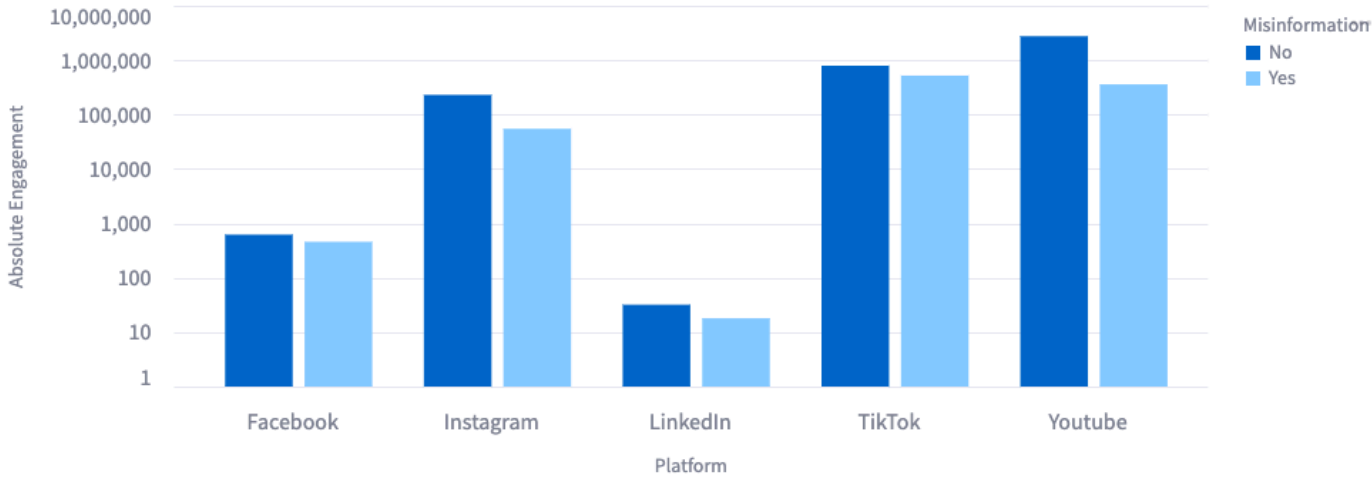
Discoverability by Country / Measurement



2a. Absolute Engagement

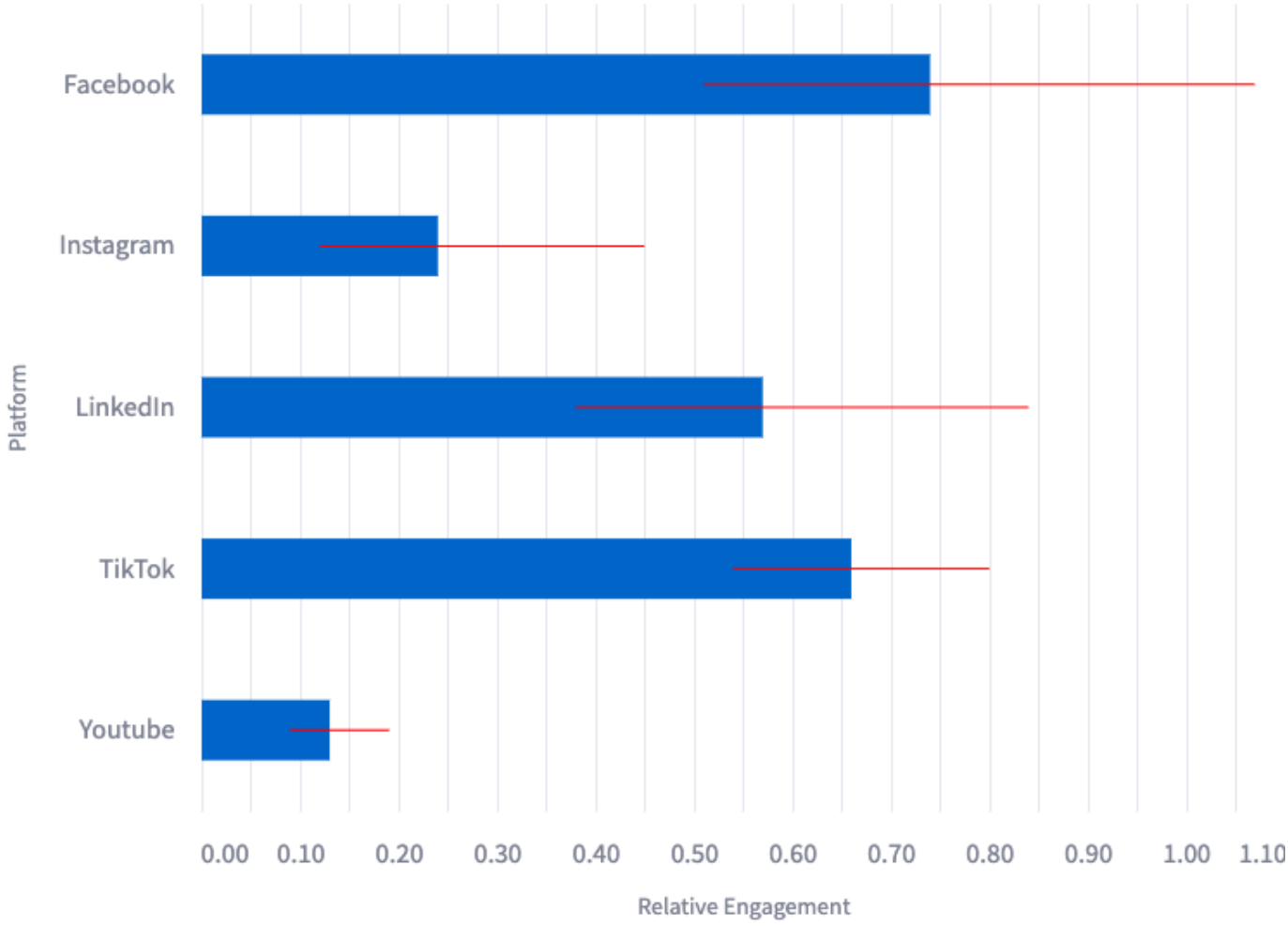
Absolute engagement refers to the magnitude, in absolute terms, of engagement with mis/disinformation content (with the caveat that the underlying data availability and nature across platforms can affect the magnitude of the metric).

Note: This chart is in log scale to allow for comparison between platforms.



2b. Relative Engagement

Relative engagement refers to the ratio of mis/disinformation engagement (where the underlying content is mis/disinformation) to non-mis/disinformation engagement (where the underlying content is non-mis/disinformation content).



Relative Engagement Components - Support

This table shows what engagement component metrics were collected across each platform.

Platform	Comments	Reactions	Shares	Views
LinkedIn	☑	☑	☑	☐
Facebook	☑	☑	☑	☐
Instagram	☑	☑	☐	☑
TikTok	☑	☑	☑	☑
Youtube	☐	☐	☐	☑

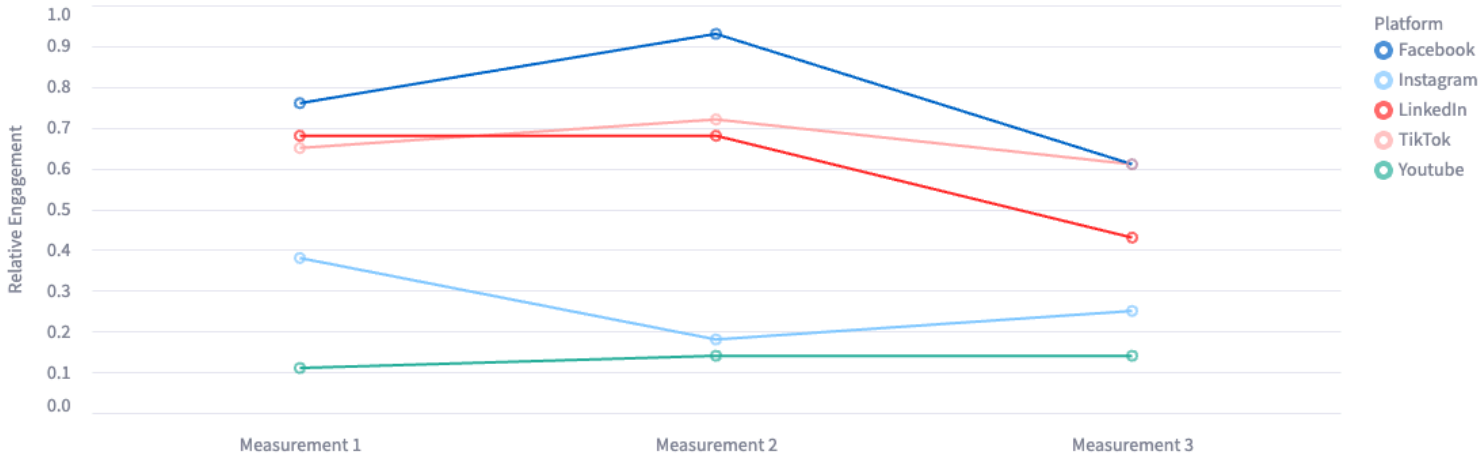
Relative Engagement Components - Scores

Overall relative engagement scores are calculated by summing all available misinformation engagement data across all components (comments + shares + reactions + views), divided by the sum of all normal engagement data across all components, sliced by platform.

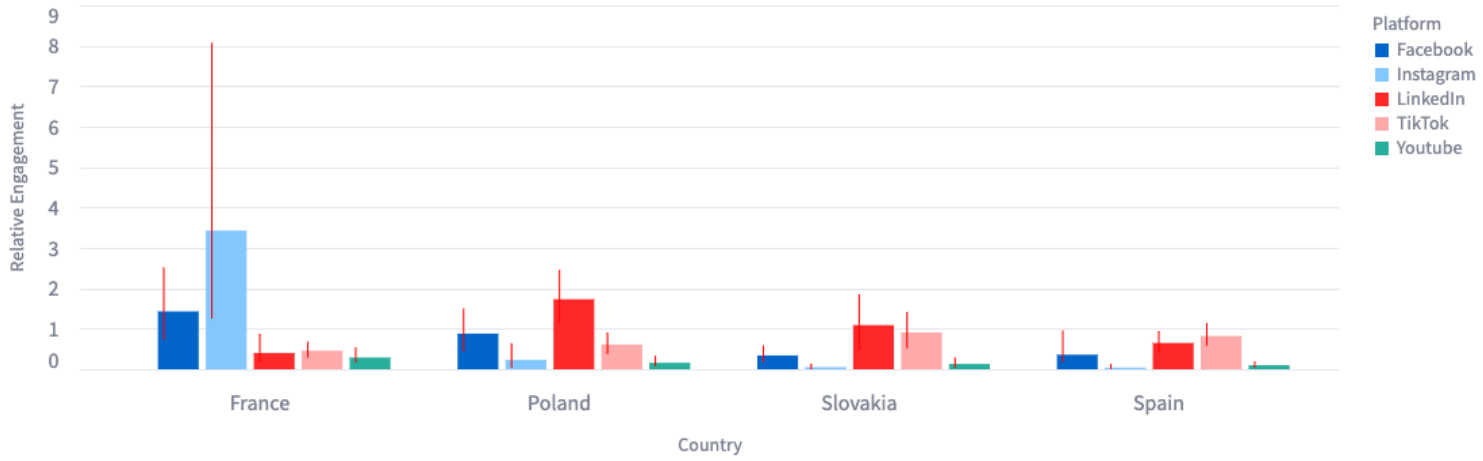
Platform	Comments	Reactions	Shares	Views	Overall
TikTok	1.06	0.49	1.04	0.66	0.66
Youtube	0	0	0	0.13	0.13
Facebook	0.41	0.56	1.58	0	0.74
LinkedIn	1.17	0.49	1.2	0	0.57
Instagram	0.33	0.25	0	0.24	0.24

2c. Relative Engagement by Measurement and Country

Relative Engagement by Measurement



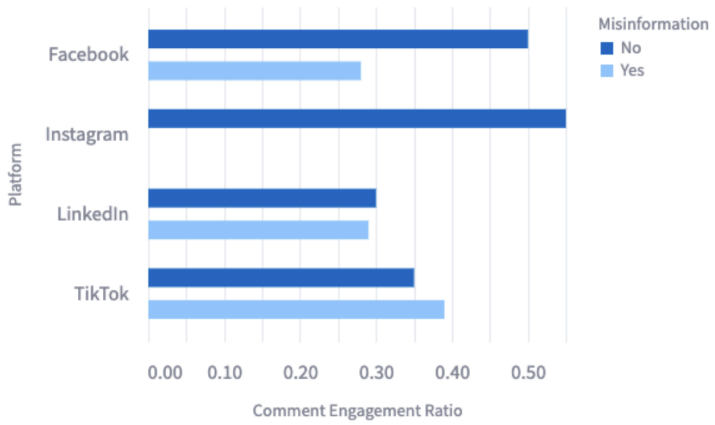
Relative Engagement by Country



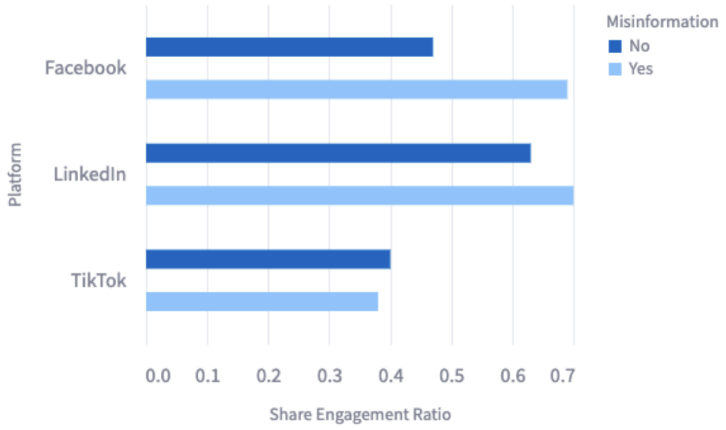
2d. Engagement Share of Top Posts

The following charts show how much engagement is driven by the top 1% of posts relative to all engagement on the platform.

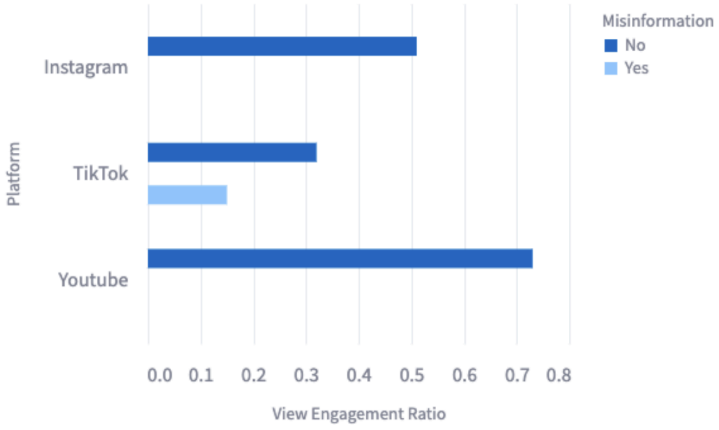
Comments: Top 1% ratio



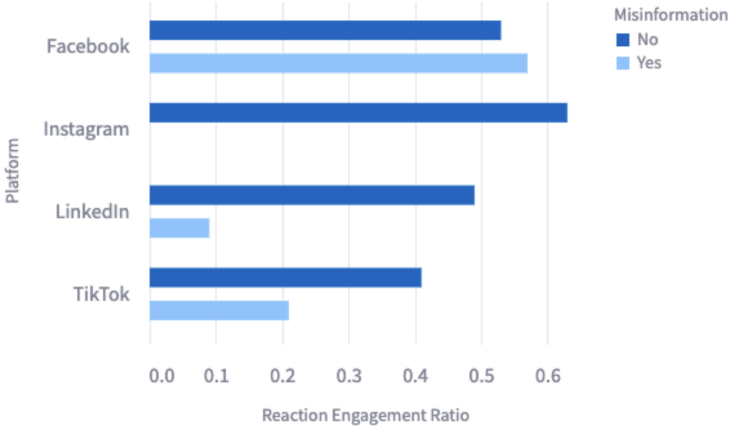
Shares: Top 1% ratio



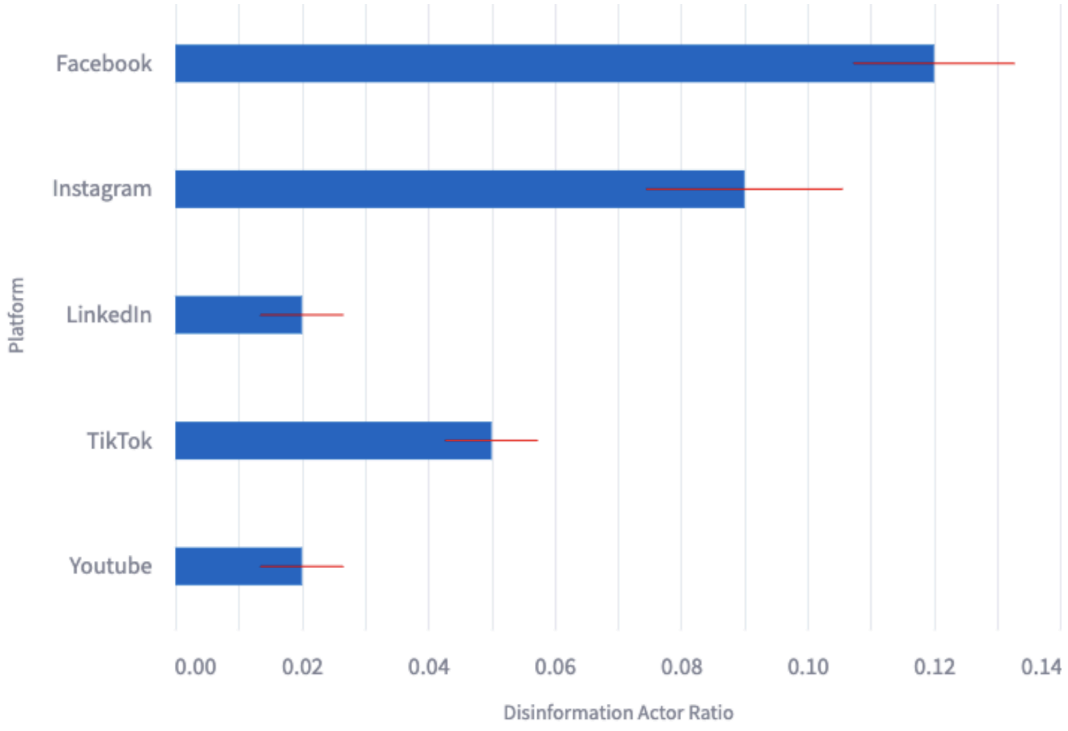
Views: Top 1% ratio



Reactions: Top 1% ratio



3a. Disinformation Actors by Platform



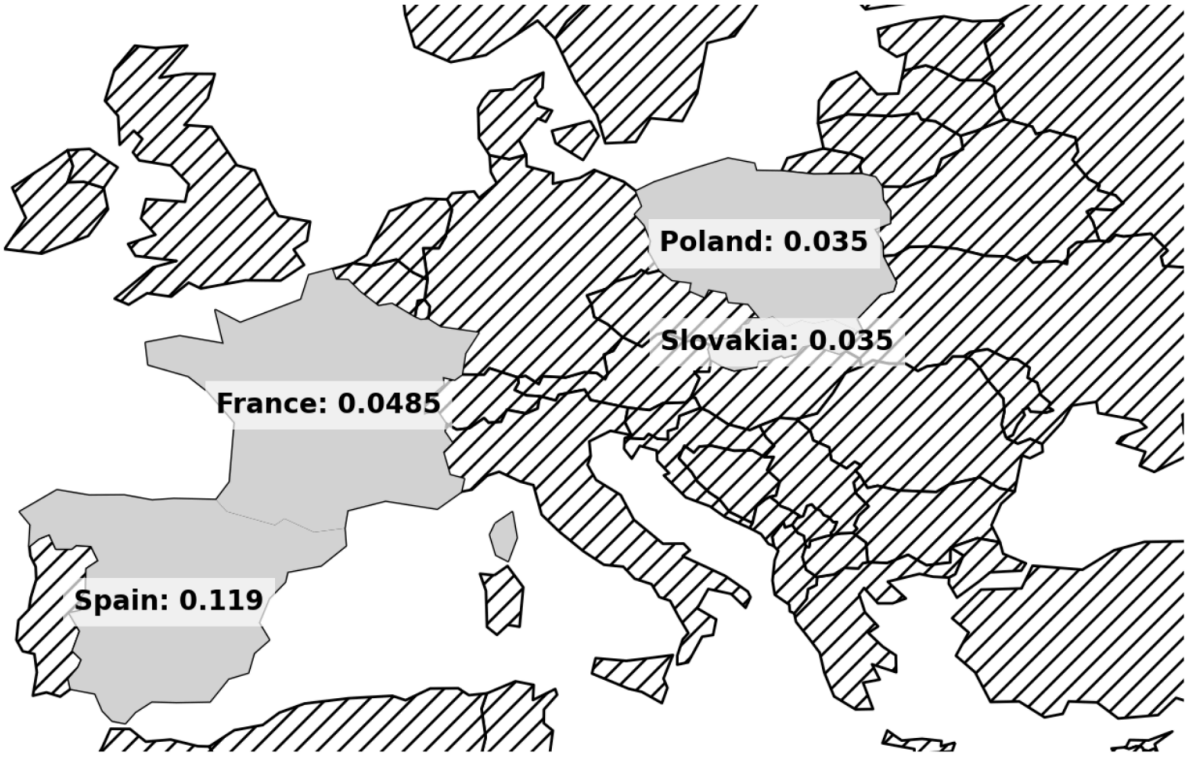
Platform	Unlikely	Low	Likely	High	Verified
LinkedIn	696	35	25	26	5
TikTok	1489	115	112	164	19
Youtube	858	39	27	21	7
Facebook	1097	190	235	244	44
Instagram	531	44	53	112	10

Accounts across all measurements were evaluated against the "ABCDE" Framework developed by the Carnegie Endowment for International Peace.

- Unlikely: The account meets a combination of 2 criteria.
- Low: The account meets a combination of 3 criteria.
- Likely: The account meets a combination of 4 criteria.
- High: The account meets a combination of 5+ criteria.
- Verified: The account is confirmed by a reputable third-party source to be actively involved in disinformation activities.

Disinformation Actors	Total Users	Disinformation Actor Ratio
652	**10,654**	**6.12%**
The sum of all verified and high probability disinfo actors.		

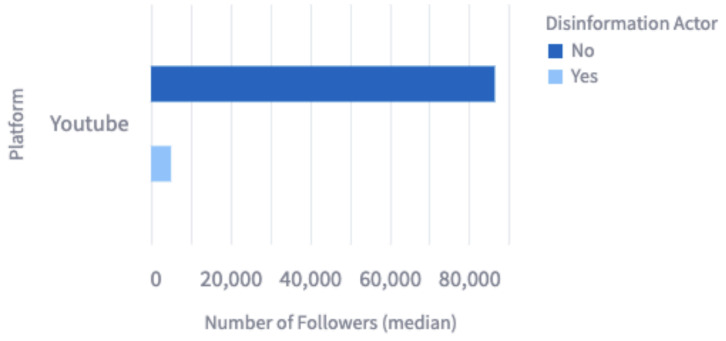
3b. Disinformation Actors by Country



Country	Disinformation Accounts	Total Accounts
Spain	351	2949
Poland	89	2546
Slovakia	97	2773
France	123	2534

3c. Disinformation Actor Engagement YouTube

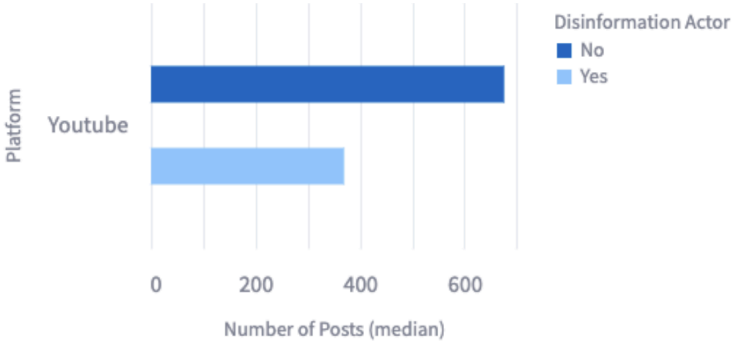
Followers



Following

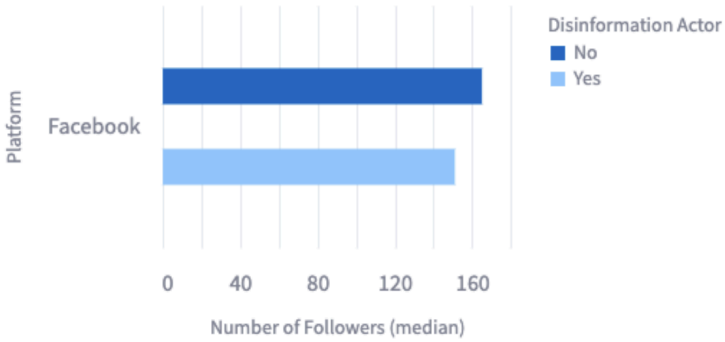
Friends

Posts

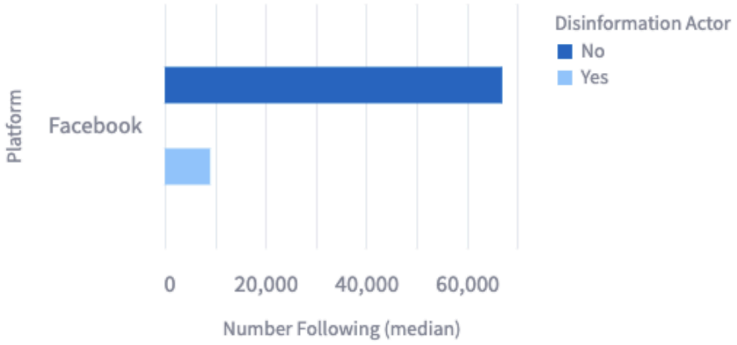


3d. Disinformation Actor Engagement Facebook

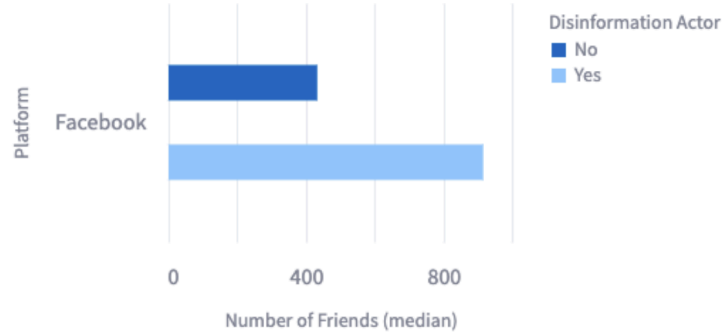
Followers



Following



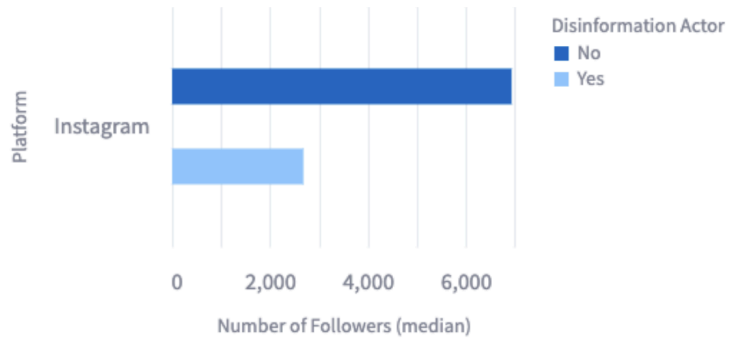
Friends



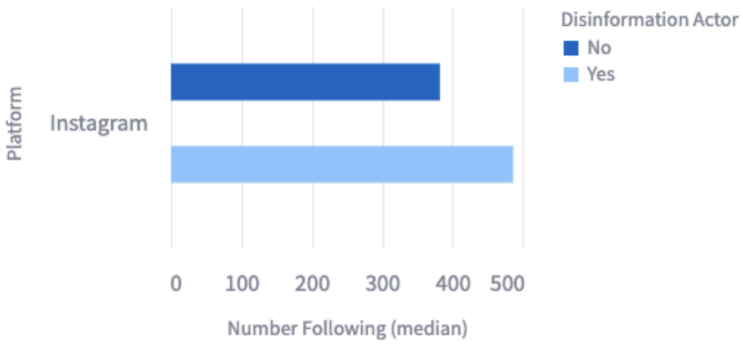
Posts

3e. Disinformation Actor Engagement Instagram

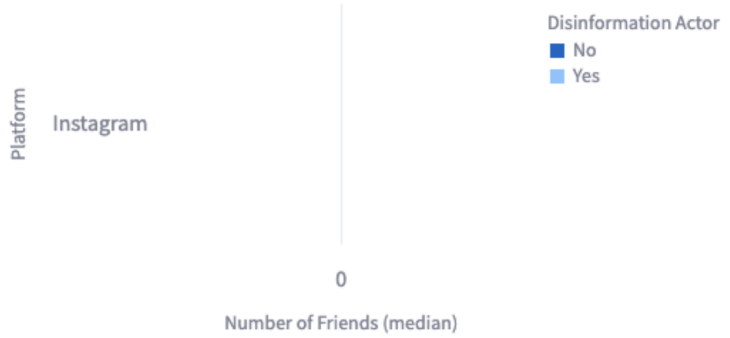
Followers



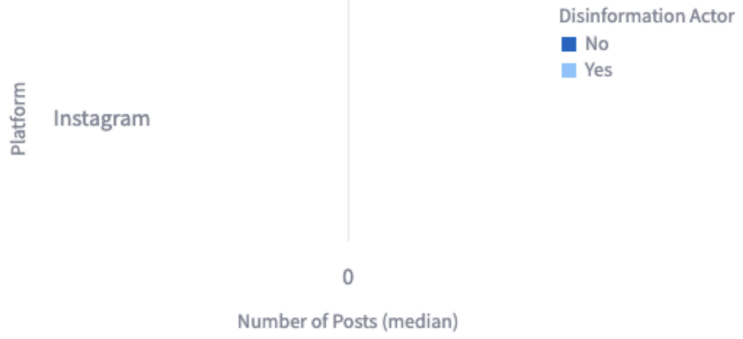
Following



Friends

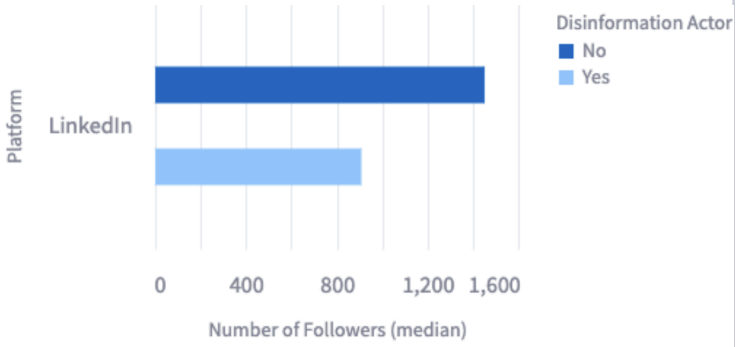


Posts



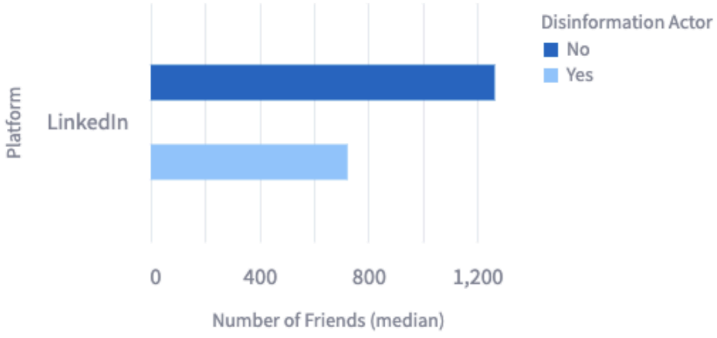
3f. Disinformation Actor Engagement LinkedIn

Followers



Following

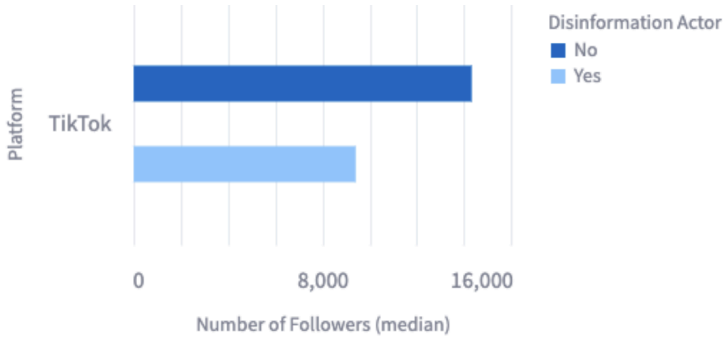
Friends



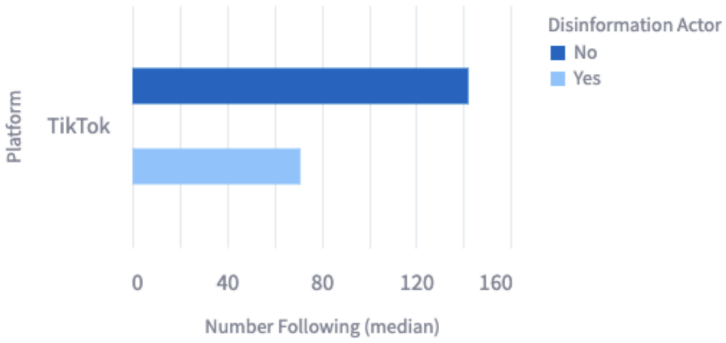
Posts

3g. Disinformation Actor Engagement TikTok

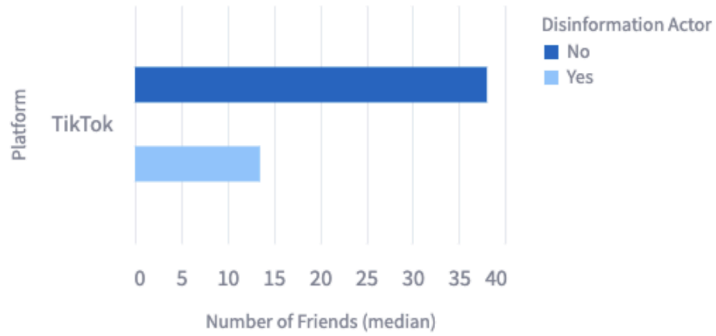
Followers



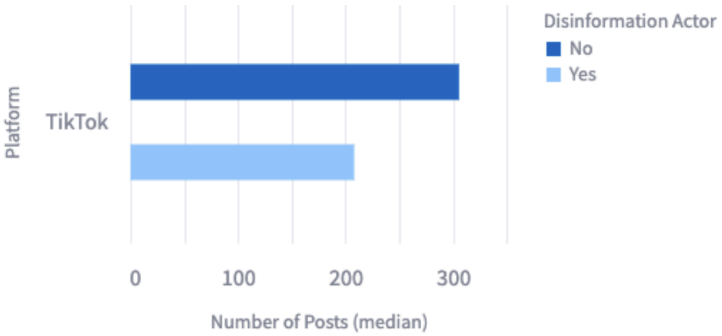
Following



Friends



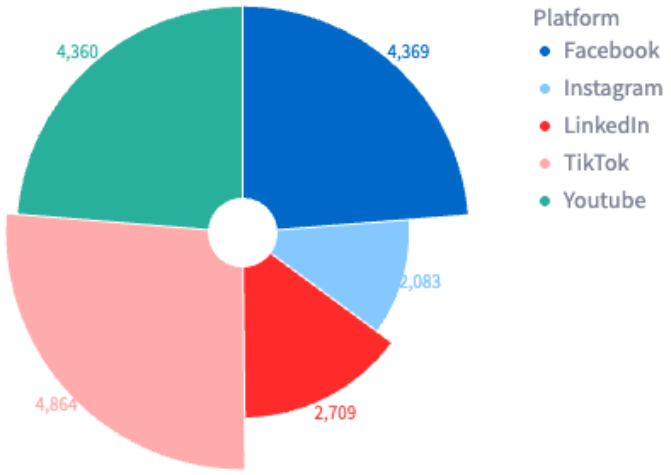
Posts



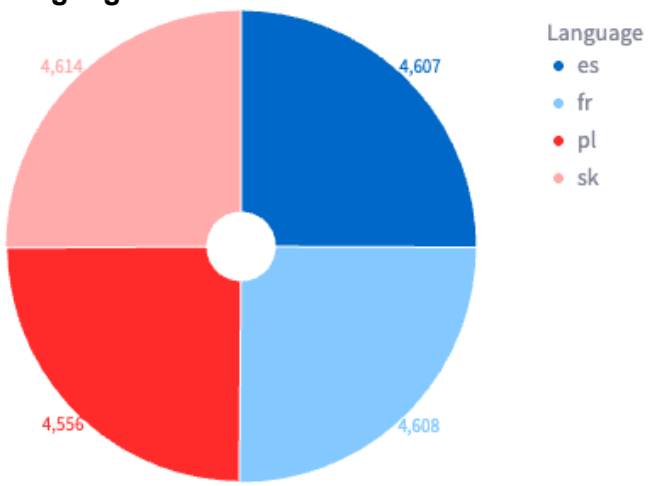
4a. Further Information

Total Posts	Misinformation Posts	Overall Discoverability Rate
18,385	**3,985**	**21.68%**

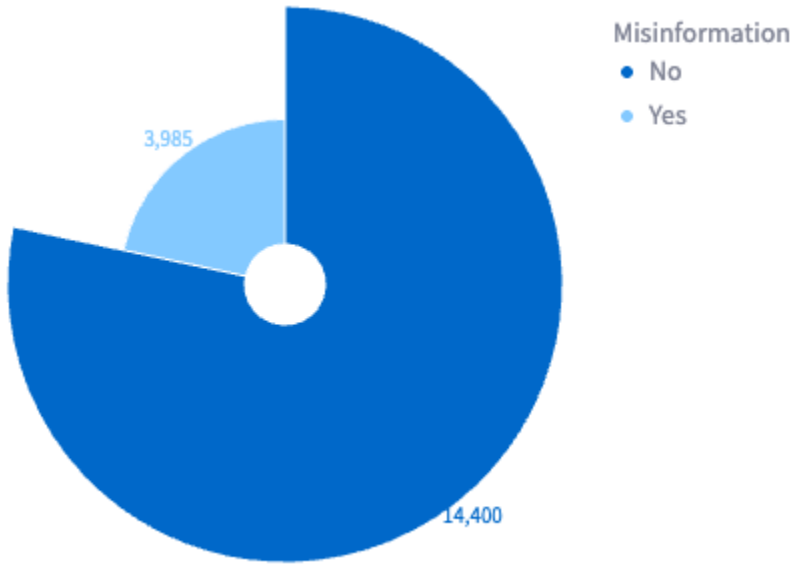
Platform



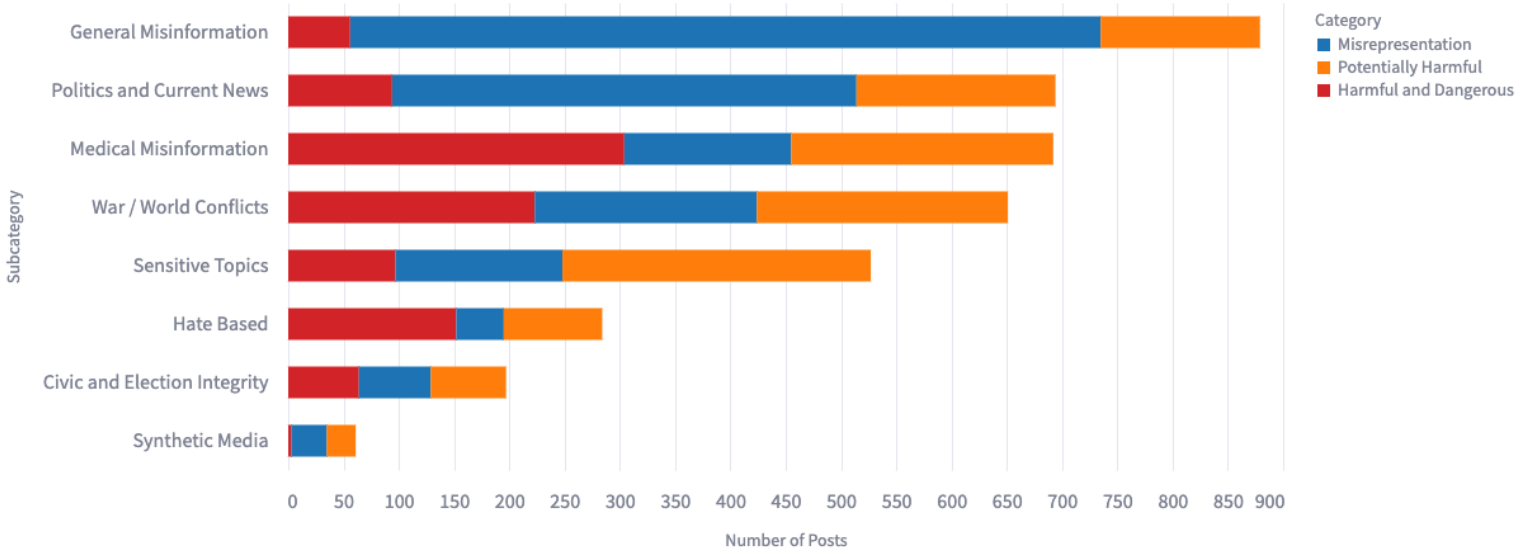
Language



Misinformation

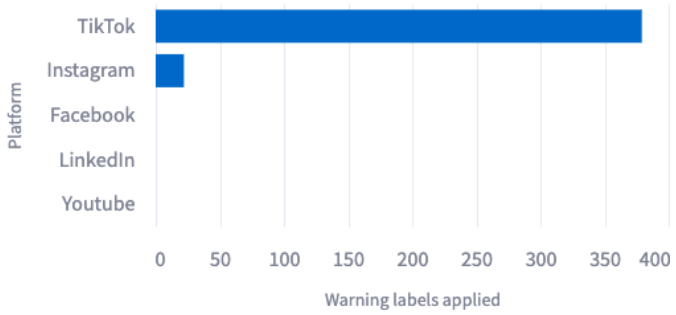


4b. Misinformation Categories

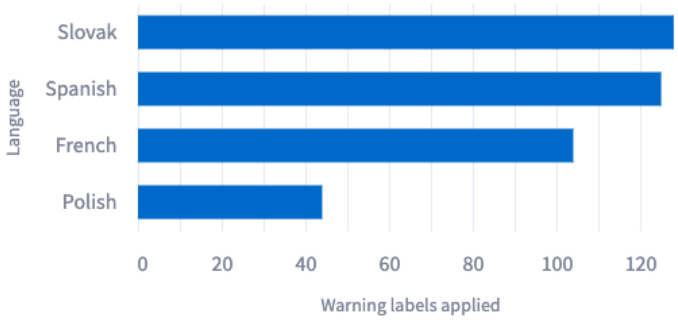


4c. Warning Labels

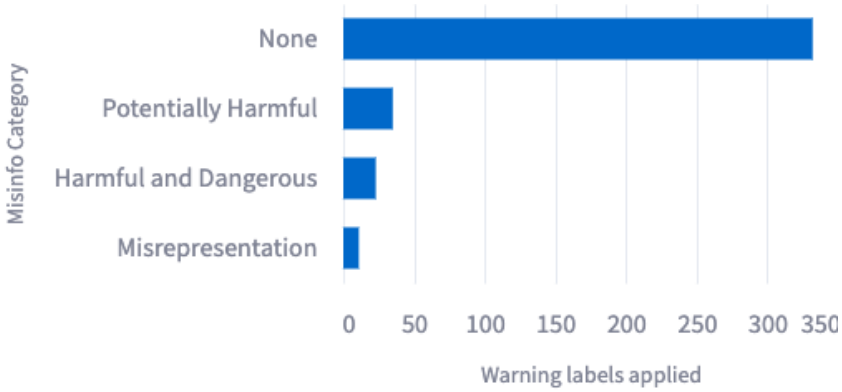
Platform



Language



Misinformation Category

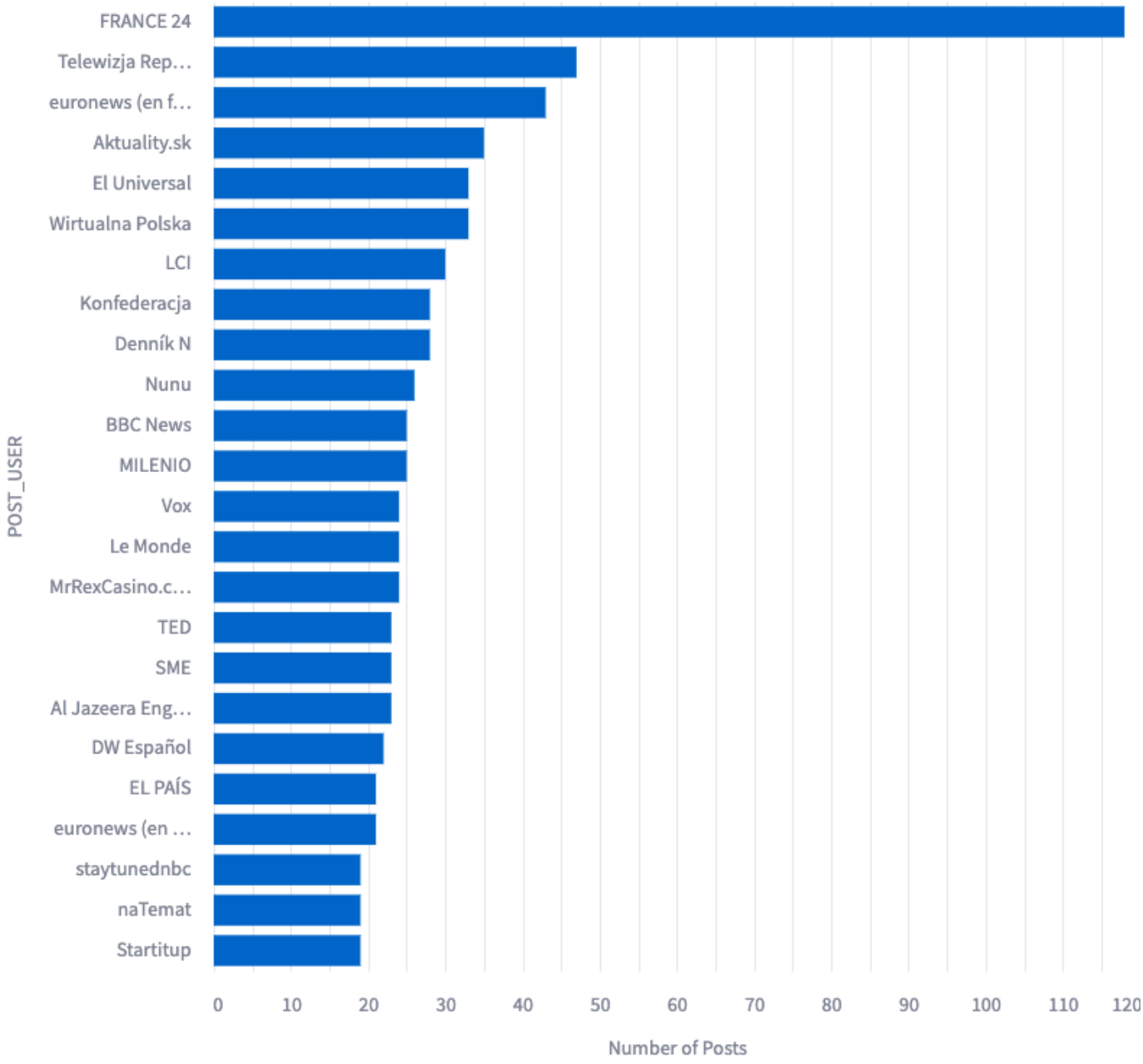


Platform	Warning Label	Total Posts
TikTok	Sensitive content	200
TikTok	Learn more about COVID-19 vaccines	110
TikTok	Learn about making informed financial decisions	31
TikTok	The actions in this video are performed by professionals or supervised by professionals. Do not attempt.	21
TikTok	Participating in this activity could result in you or others getting hurt.	14
Instagram	Missing Context. The same information was reviewed by independent fact-checkers in another post.	9
Instagram	False Information: The same information was reviewed by independent fact-checkers in another post.	5
Instagram	Sensitive Content: This video may contain graphic or violent content.	3
Instagram	False information.: Reviewed by independent fact-checkers.	3
TikTok	Get info on the UK General Election	3
Instagram	Missing context. The same information was reviewed by independent fact-checkers in another post.	2

4d. Top Users and Keywords

Top Users

Shows the top users in our sample by volume of posts returned. *This does not mean that this user is a disinformation actor*, and merely shows which users were most prevelant in our keyword searches.



Top Keywords

Shows the top keywords by volume of content returned. A distinction is shown between misinformation and non-misinformation content, giving an idea of individual keyword efficacy.