Global Disinformation Index

info@disinformationindex.org

The Global Disinformation Index (GDI) is a not-for-profit organisation focused on defunding and disrupting disinformation. We welcome the opportunity to submit the following report in response to the Carnegie Endowment for International Peace and Princeton University's inquiry on how monetisation systems shape online information environments, and what questions and data would shed further light on the financial incentives motivating disinformation threat actors.

GDI views disinformation through a lens of adversarial narrative conflict which creates division and anger among individuals and seeks to uproot trust in institutions. Since adversarial narratives exploit already existing societal tensions, **disinformation disproportionately targets marginalised groups** — including women, minorities, people of colour, the LGTBQ+ community, and other at-risk populations.

The attention economy of platforms is driven by ad dollars — the more eyeballs on a piece of media, the more profit from ad money it generates. Content that inspires strong negative emotions tends to gather the most views, which provides an economic incentive to peddle outrageous and divisive content. Similar incentives have influenced news organisations, which have relied upon digital listening analytics to determine coverage by prioritising stories that garner the most views. This leads toward unequal coverage that privileges attention-seeking behaviour; For example, as a result of this bias in the 2016 Republican primary Donald Trump received an estimated $1,898,000,000 worth of free media attention that likely had a strong influence in the outcome of that election.

- **GDI estimates that disinformation sites generate quarter billion dollars per year in ad revenues.**
- **Our estimates are conservative,** as we took the most conservative assumptions at every step of the analysis, so **the actual money generated by these ads on disinformation sites is likely much higher than our numbers suggest.**

An enhanced understanding of online information environments and monetisation systems would empower policymakers to counter the perverse economic incentives that lead to the proliferation of online disinformation and make online communities around the world safer for all.

GDI's eight priority questions address:

1. *The financial incentive to peddle disinformation.* The role of the attention economy and monetary incentives in driving the dissemination of disinformation.
2. *The radicalisation funnel.* The intersection between monetary incentives shaping online platforms and the radicalisation of users.
3. *Vulnerable groups.* The groups and individuals that are affected by monetisation systems encouraging harmful online information environments.

TABLE 1. The eight priority questions concerning monetisation of online disinformation, and example data sets desired to answer these questions

Theme	Priority Questions	Analytics, data, and collections needed
The Financial Incentive to Peddle Disinformation	**What open web disinformation is monetised, by whom, and how much?**	• Unredacted sellers.json files from all major programmatic ad exchanges. • Transaction dollar amounts for all ad transactions. • Traffic numbers to and/or number of bid requests from each site. • Data regarding how often harmful content is monetised by multiple SSPs (supply-side platforms — advertising technology companies that specialise in managing a site's ad inventory). • Data that can provide insight into how often an article or site is still monetised through another SSP or monetisation strategy when one SSP has stopped serving ads on an article or a site.
	How much disinformation is spread via monetised or paid channels? Where? What platform (eg, YT, FB) disinformation is monetised/spread via paid ads, by whom, and how much?	• Regarding YouTube — what actual monetisation is going to what channels? If we provide a channel or list of channels, can we get all the advertisers that have appeared and how much revenue the channel owner has received? How much revenue has YouTube garnered from these transactions? • Regarding social media platforms, including Facebook and Twitter — an indication of which links/posts are promoted within groups, including private groups, and what amounts were paid and by whom for those posts to be promoted.
	How much money is generated by e-commerce and online donation services monetising harmful content? What is the internal structure for e-commerce and donation platforms for identifying and preventing the monetisation of disinformation?	• Data on how much money sites peddling disinformation, or those selling products linked to disinformation, receive from e-commerce and online donation services.

Theme	Priority Questions	Analytics, data, and collections needed
The Radicalisation Funnel	**What is a typical user's journey from initial exposure to some disinformation narrative to some radical act of violence?**	• Statistics similar to CrowdTangle-like data, but for private groups — how many exist on a given topic, what is the size of the group and its growth over time, what ads are running in those groups, what links are being shared within those groups, and what is the engagement (and even more importantly, reach) of those links? • Data documenting the role of recommender system in disinfo and extremist topics (for example, the YouTube recommender system or the Facebook News Feed algorithm) — for example, data on the frequency of extremist groups being recommended, etc., as well as the frequency of "conversion" of users who join said groups after having them recommended by the platform.
	How much of the radicalisation journey is monetised or intersects with opportunities for monetisation?	• Algorithmic transparency regarding how often recommended content or creators traffic adversarial narratives or disinformation. • Data regarding what paid advertisements for creators or content is shown to what number of users and when that promoted content contains harmful or disinforming narratives.
	Where and how much of the radicalisation funnel overlaps with monetisation incentives?	• Data regarding what factors services take into account when constructing recommender algorithms for their service. • Data documenting the design goals of services when constructing algorithms and what steps are taken to mitigate bias and harm to users.
Vulnerable Groups	**Who is being exposed to disinformation? Specifically, what groups are being exposed to the harms of disinformation, and to what extent is this fueled by monetisation systems?**	• Demographics, traffic, targets – especially at the country level – of those being exposed to harmful or disinforming narratives on algorithmic platforms. • Insight on whether there are spikes of disinformation or specific kinds of disinformation on platforms around certain time periods or significant global events. • Data regarding which adversarial narratives are being monetised by what SSPs, and how much money advertising beside each adversarial narrative placed by SSPs generate.
	What internal processes are being undertaken to prevent the algorithmic amplification of disinformation? Specifically, within attention economies where the most engaging content is advantaged, what are companies doing to ensure	• Algorithmic transparency by digital services including Very Large Online Platforms (VLOPs) regarding the data and factors that determine the media shown to users — both in terms of content selection and in priority given within feeds.

Theme	Priority Questions	Analytics, data, and collections needed
	their service isn't artificially amplifying harmful content?	• The reach of harmful content and disinformation, including as a percentage of total reach of all content on the platform. • How much harmful content is being spread on the platform, including as a percentage of total content on the platform.

The Financial Incentive to Peddle Disinformation

Without intervention, digital services provided by companies are driven to maximise their profits — and on the internet, where users' limited attention translates to profit in the form of ad dollars — those companies are often driven to create the most engaging user experience. Meanwhile, e-commerce vendors pursue higher sale figures — and may sell products on an online marketplace that are associated with adversarial narratives or hate speech. Online donation services are also marketed heavily by disinformation spreaders, often asking for money under the guise of combating the threat they market to their audience. GDI has observed this in our research, in studies on the online funding of hate groups and in our DisinfoAds reports since the start of the pandemic.

The following questions surrounding the financial incentive to peddle disinformation are among the most pressing:

1. **What open web disinformation is monetised, by whom, and for how much?**

2. **How much disinformation is spread via monetised or paid channels? Where? What on-platform (eg, YouTube, Facebook) disinformation is monetised/spread via paid ads, by whom, and for how much?**

3. **How much money is generated by e-commerce and donation platforms monetising harmful content? What is the internal structure for e-commerce and online donation services for identifying and preventing the monetisation of disinformation?**

The data required to answer these questions include granular transparency of transaction dollar amounts, traffic numbers, and bid requests from each site or service. Specifically, the following data sets would be instrumental in answering these questions:

For Question 1:

- **Unredacted sellers.json files:** The ads.txt standard is an ad industry standard administered by the Interactive Advertising Bureau (IAB) to provide mechanisms for more supply chain transparency across the ad tech ecosystem. It not only includes a standard for a publisher's ads.txt file to account for each seat that that publisher has, either direct or through a reseller relationship, on a programmatic ad exchange, but it also incorporates a parallel sellers.json file on the ad exchange itself that enumerates all the publishers who have active seats on those exchanges. Theoretically, this enables cross checking between ads.txt files and sellers.json files to validate that entries in ads.txt files are indeed legitimate. However, one element of the standard allows exchanges to redact entries in their sellers.json files in order to protect the propriety of customer relationships. As a result, Google's sellers.json, by far the largest in the industry with about 2 million entries, is overwhelmingly redacted, doing little to provide the transparency that the standard was intended to provide. Not only does this limit the supply chain transparency for ad buyers, but it also limits researchers' ability to verify relationships between ad exchanges and financially-motivated disinformation peddlers. Access to unredacted sellers.json files would help shed light on the relationships between publishers of disinformation and those who enable their monetisation.

- **Transaction Dollar Amounts:** Once a relationship between a publisher and an exchange has been established, the next question would be how much money has changed hands. This information is notoriously difficult to come by, as it is considered proprietary business information. GDI has made ballpark estimates about such figures at an aggregate scale, but having granular data on transaction amounts between exchanges and publishers would help quantify the scale and scope of the problem of monetised disinformation.

- **Traffic numbers and bid requests:** Similarly to spend data, traffic numbers and/or bid request volumes to various disinformation outlets would help identify the most impactful spreaders of disinformation. While this data is also closely held, access to it would do wonders to shed further light on who is dominating the information environment and with what narratives.

For Question 2:

- Regarding disinformation spread on YouTube specifically, a breakdown of which channels are being monetised, for how much, and how the revenue is being split between the platform and the creator would be immensely helpful. For example, GDI is creating a list of known disinforming YouTube channels, and being able to associate those channels with actual monetisation activity would shed significant light on the monetised disinformation ecosystem on YouTube, with particular interest in both how much the creators *and* the platform make in terms of ad revenue for such content.

- On social platforms such as Facebook or Twitter, data on which links or posts are being promoted within groups, especially closed or private groups, and how much was paid to promote said links or posts, would indicate which disinformation actors are paying platforms to promote their content, and how much they are paying. This would be vital information to study monetisation strategies as well as platform interests in such activities.

For Question 3:

- Data on proceeds from sales of merchandise or conversion from the solicitation of direct donations to sites peddling disinformation — this data is obviously closely held, and aggregated or anonymized data would potentially suffice, but data associated with the amounts of money being generated, and on which e-commerce and donation platforms, through the sales of merchandise and solicitation of direct donations by those peddling disinformation and hate would go a long way toward illuminating this alternative source of audience monetisation.

Ethical Considerations:

- Data privacy: Fortunately, much of the data necessary for understanding the financial incentive for creating disinformation is from already public-facing and relatively large companies and publishers — such as Google, Criteo, and Amazon. It would be necessary to ensure that traffic numbers of users cannot be used to identify specific users of a service, except in the case when users include public-facing companies, platforms, or mechants. For example, a site that traffics adversarial narratives may be identified, as well as someone who sells products affiliated with hate speech on a marketplace — but the data regarding a specific individual person on a service should be protected.

The Radicalisation Funnel

Disturbingly, the monetary incentives shaping the behaviour of online platforms can cause services to incite the radicalisation of users. A wide body of literature has found that recommender systems have played a role in radicalising users — for example, GDI has published a report finding that the 'marketing funnel' strategy of social media marketing techniques can be used to radicalise predisposed users to violence, and the Facebook whistleblower revealed numerous internal documents illustrating the amplifying nature of Meta's various recommender systems when it comes to divisive hate and disinformation.

The following questions pose the utmost concern for understanding online radicalisation:

4. **What is a typical user's journey from initial exposure to some disinformation narrative to some radical act of violence?**

5. **How much of the radicalisation journey is monetised or intersects with opportunities for monetisation?**

6. **Where and how much of the radicalisation funnel overlaps with monetisation incentives?**

Key data points that are necessary for answering these questions include service of viewership content in terms of size, growth over time, what ads are running in those groups, as well as how often recommender systems are suggesting extremist content to users. Specifically, the following data would assist in answering the above questions:

Question 4:

- Right now, CrowdTangle is probably the best tool for gleaning data on what is happening on the Facebook platform. However, CrowdTangle is limited to public groups, and it doesn't provide reach information. An ideal source of data to help answer Question 4 would be CrowdTangle data, but for closed or private groups and inclusive of reach information. Reach information is particularly important in disinformation, since many users' worldview is shaped simply by the headlines that scroll past their newsfeeds, regardless of whether they engage with them or not.

- An additional data point would be internal platform data on group signups and recommender system roles in recommending groups to users. We've seen documented examples of platforms recommender systems actively assisting recruiting into extremist groups on Facebook, for example, so similar data sets showing elsewhere this may have occurred would be extremely useful in documenting these radicalising user journeys.

Ethical Considerations:

- Data privacy of specific users must be respected when collecting these analytics when the user anaylsed is not a public facing company, publisher, or merchant. Tracking a user's path down the radicalisation funnel may necessitate 'following' the data of a certain user, but the identity of the user should not be recognisable from the data given to researchers.

Vulnerable Groups

Adversarial narratives and disinformation exploits already existing social tensions, which means that the harms of disinformation disportionately impact the most marginalised in society. Researchers need a more comprehensive understanding of the risks posed by certain online services and how the harms are distributed to better protect vulnerable groups. It must be noted that adversarial narratives threaten human rights and democracy — not only through destabilising institutions, but also by driving bigotry as a political platform against marginalised and at-risk groups. It is a grossly misrepresentative caricature to trivialise the damage done by adversarial narratives as "just hurt feelings". Disinformation puts lives directly at risk, and this harm is even more likely to materialise when adversarial narratives are uptaken by politicians which can institute legislation and abuse the power of government against targeted groups. For example, exposure to derogatory language has been found to increase political radicalisation and deteriorate intergroup relations.

These questions are of special concern to groups and individuals that are affected by monetisation systems encouraging harmful online information environments:

7. **Who is being exposed to disinformation? Specifically, what groups are being exposed to the harms of disinformation, and to what extent is this fueled by monetisation systems?**

8. **What internal processes are being undertaken to prevent the algorithmic amplification of disinformation — specifically, within attention economies where the most engaging content is advantaged, what are companies doing to ensure their service isn't artificially amplifying harmful content?**

In order to best answer these questions, we would ideally like to understand the demographics, traffic numbers, and targeting characteristics — especially at the country level — of those being exposed to harmful, toxic, or disinforming narratives on all of the algorithmic platforms. This would allow us to identify the groups that are most at risk and most likely to be affected by algorithmically amplified or monetised disinformation online, and where best to deploy resources or countermeasures to reduce the risk of a resultant harm.

Additionally, it would be useful to know whether these "walled garden" platforms have observed spikes of disinforming content that correlate with certain global, regional, or local events or within specified time periods. For example, we have observed spikes in disinforming content that correlate with new COVID vaccine announcements, or with global events like the United States rejoining the Paris Climate Accords. It would be helpful to have data that would illuminate if the same effect occurred within platforms that correspond to national elections, wars, or other such significant events.

Finally, regarding what internal processes platforms are undertaking to prevent algorithmic amplification of disinformation, general algorithmic transparency by platforms would be critical. It is vital to know what data and factors platform companies use in recommender systems, news feeds, and search results to determine which content gets selected or prioritised to users. In addition, it would be important to understand just how much content, both in absolute numbers and as a percentage of the total amount of platform content, falls under the category of harmful, toxic, or disinforming narratives. Lastly, it would be useful to know the total reach of that content, both in absolute terms and as a proportion of the total reach of all content across the platform.

Ethical Considerations:

- As with all of the data suggested in this document, the data obtained by researchers to answer the questions in this category should not be able to be used to discern the identity of a specific user of a service, in the case where the user is not a public-facing company, publisher, or mechant.

Summary and Conclusion

GDI views disinformation in large part as a destructive externality of a toxic business model, and a problem that is largely motivated by perverse financial rewards for both content creators and intermediating platforms. In order to combat the problem, comprehensive and transparent data is required to both illuminate the extent of the issue and identification of those most at risk and enable the creation and enforcement of regulation.

In this paper we've outlined eight key questions that such data must answer, and suggested a collection of data sets that would go a long way in answering them. Of course, not only are there ethical considerations to be accounted for when collecting and holding this information, which we've outlined, but there will be inevitable pushback by those commercial entities that hold these data to them being made available. This does not mean that they are not vital, nor does it mean that accessing them is impossible in light of emerging regulatory authorities requiring such transparency. Should those efforts progress, our hope is that the data sets enumerated here provide a foundation for the kinds of data that regulators around the world demand, and which are provided to researchers in a centralised way in order to facilitate the necessary oversight over and transparency into the most important information infrastructure in the world.