The Rt Hon Sajid Javid MP, and  
The Rt Hon Jeremy Wright MP  
UK Government  
Department for Digital, Culture, Media &  
Sport  

21 June 2019

Dear Minister Javid and Minister Wright:

Disinformation has been used as a tool to weaponise mass influence and disseminate propaganda. It has brought extreme fallout for economies, politics and societies around the globe. No country is immune from these online and offline harms.

We welcome the UK government’s efforts to more effectively and proactively respond to disinformation and the harms it creates. Please find attached our submission to the consultation on the “Online Harms White Paper” that has been launched.

The Global Disinformation Index (www.disinformationindex.org) is a UK-based not-for-profit whose mission is to restore trust in the media by providing real-time automated risk ratings of the world’s media sites. We are currently piloting the index and tracking disinformation efforts in the UK and globally. Some of our research and thought pieces speak directly to how best to define disinformation and the harms it creates.

We believe that to combat disinformation, we need to understand efforts to disinform – both upstream (where disinformation starts) and downstream (where and how it spreads). This is where the Global Disinformation Index (GDI) has set its focus.

We view this submission as setting out such a policy framework to address online harms.

We look forward to working with the UK government in these efforts.

Sincerely,

[Signature]

Clare Melford  
Executive Director and Co-Founder  
The Global Disinformation Index
**RECOMMENDATIONS:** The Global Disinformation Index (GDI) believes there are a series of recommendations for ensuring that regulating online harms can be done “in a targeted and proportionate manner” (question 8 of the consultation). These include:

1. **Specify more clearly the types and sources of harms to be covered by regulation.**
   - Rationale: Any regulatory regime which penalises platforms for not removing a certain type of content and fails to clearly define and limit the scope of affected content is very likely to lead to platforms excessively removing content in that category in order to minimise risk.

2. **Define and frame disinformation more specifically as a clear source of online harms.**
   - Rationale: For the GDI, to remedy disinformation, one has to look more specifically at the actions behind it: the verb - to disinform. We define ‘to disinform’ as: to purposely and/or maliciously mislead by spreading inaccurate information (in terms of the content itself and the context). The whitepaper does make a basic distinction between disinformation and misinformation. However, disinformation should not just be about the content but the context it is presented - and the narratives, networks and actors behind it.

3. **Define the range of harms arising from disinformation.**
   - Rationale: The Online Harms Whitepaper places disinformation within the category of ‘harms with a less clear definition’ (Table 1), alongside items such as trolling and extremist or violent content. However the paper offers few other details for understanding or addressing the harms posed by the spread of false and malicious information. The harms identified in the white paper should be re-conceptualised instead as ‘harmful activities,’ and a typology of harms should be established. For the GDI, this should be based on the substance of the harm inflicted, such as:
     1. Increased social polarization
     2. Fraudulent/faulty basis for decision-making
     3. Erosion of trust in institutions
     4. Reputational harms to persons or groups and harassment
     5. Psychological distress and threats to personal safety
     6. Discrimination (e.g., hate speech)
     7. Incitement to violence
     8. Administrative burdens to affected persons (to correct the harm)

4. **Address remedies to combat disinformation that do not target specifically content removals.**
   - Rationale: We very much welcome the government’s stated interest in not becoming an arbiter of the truth (Executive Summary, §38) or directly resolving content moderation disputes. However, these goals are at odds with the government’s desire to hold companies accountable to their terms and conditions (§5.2). By identifying the harm as the disinformation itself, the white paper places emphasis on regulating content. This is likely to lead to disputes between companies and the regulator over how false content is to be moderated (even despite the government’s best intentions to avoid it).
5. Understand and regulate the process of disinforming.
   ● Rationale: The GDI believes that the best way to combat disinformation is to address the actions, actors and activities that allow disinformation to spread - rather than the specific content. Such an approach would ensure that the UK regulator does not become involved in content moderation disputes. It would help to protect the regulator from having disputes arising from claims of suppressing political or other types of thoughts and beliefs. It would also help to rationalise its resources and capacities. If an approach to regulate content was pursued, it would require the regulator to become skilled about the standards and practices of each individual platform (given the complexities and specificities of content moderation as well as different community standards). The process of disinforming is facilitated by many actors that do not host or produce disinformation themselves. For example, the GDI demonstrates in a forthcoming report how the ad tech industry funnels money to “junk” domains that are known to disinform their readers, thereby making the process more profitable.

6. Adopt a cross-platform approach.
   ● Rationale: The GDI fully believes that the UK regulator should recognise and remedy the cross-platform process of producing harmful content. From our network mapping of disinformation actors (e.g. those promoting the QAnon and anti-5G narratives), harmful actors often disperse their activity across a range of platforms - using different services to recruit, organise, monetise, and harm others. Mitigating harm on one platform will often require that the regulator takes a broader view of the extent to which that harmful activity moves between services.

7. Develop standardised metrics to assess harms.
   ● Rationale: The companies subject to regulation are diverse - they offer a range of services, are various sizes, and suffer from different vulnerabilities. The regulator will need to develop standardised metrics for comparing the degree of harm posed by these different services, to determine risk and set priorities for intervention. The proposed regulatory approach can be improved by developing a clearer typology of harm so that the regulator can identify harms based on their consequences, rather than the activities that produce them (see Recommendation 3).

8. Recognise and address “bad-faith” complaints.
   ● Rationale: There seems to be an unstated assumption in the white paper that user complaints regarding content moderation decisions or other alleged terms and conditions violations are always in good faith. This position does not correctly capture the adversarial landscape navigated daily by platform content moderators who do this work. Indeed, the mass flagging of a user’s content is a common tactic harassers use to silence their targets. There should be a set of clear public criteria for distinguishing what constitutes good and bad faith complaints and appeals, which is rigorously applied.