CENTRE FOR EUROPEAN REFORM
LONDON ★ BRUSSELS ★ BERLIN

Insight

Will the Digital Services Act save Europe from disinformation?

by Zach Meyers, 21 April 2022

In negotiating the Digital Services Act, EU law-makers balanced tackling disinformation with protecting free speech. The Commission's last-minute proposal for stricter regulation of tech platforms during crises undermines this balance.

Online disinformation – material propagated with the intention to mislead – is a serious threat to the EU. It has contributed to many of the EU's recent challenges: panic about immigration, the rise of the far right and left, Islamophobia, vaccine hesitancy and Brexit. Europe's rivals, in particular Russia and China, use disinformation campaigns as low-cost and low-risk methods to foment dissent and promote their preferred narratives about these issues in the EU. Russia's invasion of Ukraine is the latest battleground online. Platforms like Twitter, YouTube, TikTok and Instagram have been flooded with Putin's lies about the 'Nazification' of Ukraine.

This flood of disinformation comes as the EU is finalising the Digital Services Act (DSA), a major new law designed to regulate online platforms, including social media platforms like Facebook, Twitter and TikTok which are used to disseminate disinformation. The DSA forces large platforms to be more transparent and accountable for tackling disinformation. As law-makers finalise the DSA, the European Commission has begun insisting it needs stronger powers to direct how platforms tackle disinformation during crises. These powers would undermine the careful compromises law-makers have already agreed in the DSA – and risk making platforms' responses to disinformation worse.

In the EU, spreading false or misleading information is not generally illegal. Freedom of expression includes the right to express incorrect views. And the distinction between 'fake news' and 'legitimate opinion' is often contested. Despite the EU's recent decision to ban Russian media outlets Russia Today and Sputnik from broadcasting in the EU, policy-makers generally recognise that simply banning disinformation is not a realistic or desirable option. Instead, the EU has sought to curb the impact of lies peddled online in ways which preserve free speech. For example, the EU's 2018 Action Plan against Disinformation focused on identifying disinformation, supporting independent media and fact-checkers, and promoting media literacy. The EU's European External Action Service (EEAS) also set up strategic communications divisions, known as the StratCom Task Forces. The EU's 2020 European Democracy Action Plan also established a framework for collecting evidence about foreign disinformation campaigns. As the Kremlin propagated lies about its invasion of Ukraine, for example, this evidence allowed the EU High Representative for Foreign Affairs and Security Policy to quickly name these strategies publicly and correct false information.

Disinformation is hardly limited to the online world. More polarised people are less likely to use social media – being older, they tend to rely on newspapers and TV. Conversely, most users of social media are exposed to a wide diversity of opinions. But – perhaps because regulating foreign tech firms is easier than tackling problems with some of the EU's own media outlets – lawmakers remain focused on the importance of online platforms. Though they each have different acceptable use policies, online platforms do not typically ban all disinformation, both because identifying misleading material is difficult and to protect freedom of speech. More important is how disinformation is amplified. Rather than showing a chronological view of posts, most platforms now use personalised algorithms, designed to show users the most relevant and engaging content. Disinformation is often designed to exploit these algorithms, being emotive to attract user engagement. 'Troll factories', like Russia's so-called Internet Research Agency, also co-ordinate the actions of many different user accounts, tricking algorithms into believing that content is genuinely engaging, so that platforms show that content to more users.

Currently, the EU primarily relies on platforms self-regulating to avoid these problems. Self-regulation focuses on easier issues, like regulating online advertising and increasing the prominence of reputable news sources. But even though the EU is trying to strengthen self-regulation, voluntary steps will probably remain insufficient. For example, self-regulation has not forced online platforms to devote enough resources to protecting EU users. Facebook chose to deploy more content moderators in other parts of the world than EU member-states, particularly the US, Brazil and India. Disinformation in the largest EU languages, including Italian, French, Portuguese and Spanish, is far less likely to be quickly assessed than content in English. This is a concern because many disinformation campaigns are deployed rapidly and locally. For example, in recent weeks, Russia has been particularly intent on stoking anti-Ukrainian sentiment in eastern EU member-states such as Poland.

If it is passed by EU law-makers, the DSA would shift away from self-regulation in relation to disinformation. Regulation would, however, be light-touch. Platforms' acceptable use policies would need to be clear and diligently enforced. Platforms would also need to be more transparent about their content moderation. For example, they would need to publish annual reports explaining their content moderation practices, including how many content moderators they allocate in each EU member-state official language. They would also need to put in place a complaints mechanism, so that users can challenge a platform's decision to either remove content or to ignore a complaint about the content. The very largest platforms would have to comply with somewhat stricter rules. They would need to formally assess some of the risks which stem from the design and use of their services, such as risks of their services being intentionally manipulated to impact civic discourse. Large platforms would also have to explain how they mitigate these types of risks. They would also be encouraged to participate in 'crisis protocols', which would be drawn up by the Commission and would explain how platforms should deal with extraordinary events like wars, where online platforms could be misused to spread disinformation.

The DSA has disappointed some commentators, because until recently negotiators agreed that platforms should be free to choose how they should handle disinformation. The agreed approach was that platforms could decide whether to allow 'lawful but awful' content, such as pro-Russian propaganda, and – if they allow it – how to mitigate its impact. For example, platforms may decide to remove such disinformation. Or they may label it with warnings, provide fact-checked information with it, deamplify or stop recommending it, or provide subsequent facts to users who have been exposed to it. This reflected a careful balance: many MEPs and member-states are rightly resistant to regulate lawful content, and preferred to focus on ensuring platforms were transparent and accountable for how lawful content was handled.

However, in a last-minute change, the Commission has complained that the "anticipatory or voluntary nature" of the obligations to tackle disinformation would be insufficient in a 'crisis'. The Commission argues it must be able to direct how platforms respond to crises like the Russian invasion of Ukraine, and the Commission wants the power to determine whether there is such a 'crisis' itself. This proposal undermines the careful balance between MEPs and member-states and risks making platforms' response to disinformation worse.

If it had these powers, the Commission would undoubtedly feel pressured to force large platforms to simply remove pro-Russian 'fake news' – similarly to how the Commission banned Russia Today and Sputnik. However, requiring systemic removal of such information would inevitably have to rely on machine-learning tools, which are notoriously inaccurate, fail to have regard to context, and therefore often impact important, genuine content – such as parody and legitimate reporting. Most recently, for example, Twitter mistakenly removed accounts which were providing a valuable service by reporting on Russian military activities in Ukraine. An emphasis on large-scale removal of harmful material is also likely to prompt users to flee to smaller and less scrupulous platforms. This explains why some online platforms are selective about the types of harmful content they disallow. For example, Facebook only removes material that could cause imminent physical harm; that comprises vaccine misinformation; that interferes with elections; or 'deep fake' videos. Even if the Commission directed platforms to take less onerous and seemingly sensible steps to tackle disinformation, those steps might also have unintended negative consequences. For example, when platforms attach 'warning' labels to disinformation, users may wrongly assume everything unlabelled is trustworthy. This is a problem when content relates to current affairs, where independent fact-checkers have not yet determined what is true.

In comparison, steps which may superficially look like 'weak' responses to disinformation – such as asking a user to confirm that they want to share unverified news, before allowing them to do so – can have an outsized impact, dramatically reducing the propagation of fake news. Researchers, policy-makers and platforms themselves are still working out which responses by platforms are most effective. Studies still point in conflicting directions. Yet, in a crisis like Russia's invasion of Ukraine, the Commission is unlikely to be satisfied with steps which look weak, even if these are the most effective options.

Responding to disinformation by fiat is also unlikely to be effective because disinformation tactics change rapidly – as do the required responses. For example, co-ordinated disinformation campaigns now tend to be more localised and targeted; they increasingly try to engage influential individuals to propagate their messages; and they use tools to make co-ordination harder to detect.

EU law-makers could learn from the UK government's experience drafting the Online Safety Bill, Westminster's equivalent of the DSA, which has just been tabled in parliament. The UK government's original proposal would have empowered a public regulator, Ofcom, to direct platforms about how to deal with harmful content such as disinformation. Yet after widespread criticism that this would cripple freedom of expression, the government swung to the opposite extreme: it has abandoned any attempt to regulate socially harmful material like disinformation. Instead, the bill would now only regulate how platforms tackle content which is illegal or which is lawful but poses 'physical or psychological harm' to individuals. EU law-makers have negotiated a more moderate and balanced approach to disinformation, which has avoided much of the criticism and controversy in the UK. Yet the Commission's last-minute proposal for "crisis protocols" would undermine this approach.

Despite years of efforts, the EU remains vulnerable to disinformation and misinformation campaigns, which threaten to polarise societies and reduce trust in democratic institutions. Given the sensitive interests at stake, such as freedom of speech, EU lawmakers have done well to agree on new responsibilities for online platforms. Other Western countries are far behind in achieving any consensus. The UK's Online Safety Bill is still attracting criticism even amongst Tories. And US Congress agrees on the need to reform the current law regulating online platforms but is hopelessly divided on what form that regulation should take. EU law-makers have settled on a good plan for mitigating the impacts of disinformation online – by making platforms more transparent and accountable. The Commission's last-minute effort for tougher crisis powers could undermine the compromises already achieved, stifle free speech, and risk making disinformation worse.

Zach Meyers is a senior research fellow at the Centre for European Reform.

A number of technology companies including Amazon, Apple and Facebook are corporate members of the CER. The views expressed here, however, are solely the author's, and should not be taken to represent the views of those companies.