

DEMOS

DRIVERS OF DIGITAL DISCORD

HOW NEWS MEDIA AND
SOCIAL MEDIA DRIVE ONLINE
DISCOURSE - AND PATHWAYS
FOR CHANGE

ELLEN JUDSON
SOPHIA KNIGHT
KIERAN YOUNG
SHAUN RING

DECEMBER 2023

Open Access. Some rights reserved.

Open Access. Some rights reserved. As the publisher of this work, Demos wants to encourage the circulation of our work as widely as possible while retaining the copyright. We therefore have an open access policy which enables anyone to access our content online without charge. Anyone can download, save, perform or distribute this work in any format, including translation, without written permission. This is subject to the terms of the Creative Commons By Share Alike licence. The main conditions are:

- Demos and the author(s) are credited including our web address **www.demos.co.uk**
- If you use our work, you share the results under a similar licence

A full copy of the licence can be found at **<https://creativecommons.org/licenses/by-sa/3.0/legalcode>**

You are welcome to ask for permission to use this work for purposes other than those covered by the licence. Demos gratefully acknowledges the work of Creative Commons in inspiring our approach to copyright. To find out more go to **www.creativecommons.org**



All content and editorial decisions were made exclusively by Demos throughout research and report development. This independent research was supported by The Archewell Foundation.

Published by Demos December 2023
© Demos. Some rights reserved.
15 Whitehall, London, SW1A 2DD
T: 020 3878 3955
hello@demos.co.uk
www.demos.co.uk

CONTENTS

ACKNOWLEDGEMENTS	PAGE 4
EXECUTIVE SUMMARY	PAGE 6
GLOSSARY	PAGE 9
INTRODUCTION	PAGE 10
PART ONE: HOW NEWS MEDIA, SOCIAL MEDIA AND CITIZENS INTERACT TO DRIVE ONLINE HARMS	PAGE 13
PART TWO: WHAT WOULD A HEALTHY INFORMATION ENVIRONMENT LOOK LIKE AND WHAT'S PREVENTING IT?	PAGE 47
PART THREE: POLICY RECOMMENDATIONS AND A PATHWAY FOR CHANGE	PAGE 55
CONCLUSION	PAGE 63
ANNEX ONE: METHODOLOGY	PAGE 64

ACKNOWLEDGEMENTS

Huge thanks are owed to all those who made this work possible.

Thanks to the team at the University of Sussex, including David Weir, Albertus Andito and Andrew Robertson, for their support in designing and leading the computational data analysis. We are grateful to the team at Purpose for their support and exchange of ideas through their concurrent research.¹

Thanks also goes to the wider Demos team, including Felix Arbenz-Caines, Chloe Burke, Lucy Bush, Polly Curtis, Alice Dawson, Hannah Perry, and Fellows Oliver Marsh and James Ball for their contributions to the project, as well as former colleagues Alex Krasodonski-Jones and Josh Smith.

All those who contributed to our research, including the representatives from government, regulators, including DSIT and Ofcom, as well as news media and civil society who participated in our policy workshop, and those who participated in interviews to share their insights.

- James Ball, journalist, author and Demos Fellow
- Professor Charlie Beckett, Director of Polis and the Polis/LSE JournalismAI project
- Guilherme Canela, Chief of Freedom of Expression and Safety of Journalists at UNESCO
- Matthew d’Ancona, Editor-at-large, The New European and Contributing Editor, Prospect Magazine
- Charlotte Dewar, Chief Executive of the Independent Press Standards Organisation (IPSO)
- Jonathan Heawood, Executive Director of the Public Interest News Foundation
- Elliot Higgins, Founder of Bellingcat
- Professor Jeff Jarvis, Director of the Tow-Knight Center for Entrepreneurial Journalism at the City University of New York
- Lexie Kirkconnell-Kawana, Chief Executive Officer of IMPRESS
- Nic Newman, Senior Research Associate at the Reuters Institute for the Study of Journalism
- Meera Selva, Chief Executive Officer of Internews Europe
- Four other experts who contributed anonymously, including those working in public policy and advocacy for news organisations, technology companies and civil society organisations.

This report was produced by Demos and is editorially independent. As ever, any errors remain the authors’ own.

Ellen Judson
Sophia Knight

¹ <https://www.purpose.com/australian-news-media-and-online-hate-speech/>

ABOUT THIS PROJECT

Demos is Britain's leading cross-party think tank. We put people at the heart of policy-making to create bold ideas and a more collaborative democracy.

CASM, Demos' digital research hub, works to investigate, articulate and advocate for an internet and technologies that protect democratic values and human rights. This project is part of our ***Strengthening Information Environments*** programme. This programme looks at how we can build a more healthy, resilient and sustainable information environment for citizens in support of their democratic and digital rights.

EXECUTIVE SUMMARY

This project aims to explore the dynamics of news information and public discourse in online spaces. We use a mixed-methods approach to investigate how news media, social media and citizens are interacting to drive digital discourse and online harms. We also set out to explore what a healthy information environment would look like, and what structures are preventing it from being realised. Ultimately, our aim is to set out a vision for how we can achieve better information environments and healthier democratic discourse in a digital age. This is the pathway to change.

We find that our relationship with information has been uprooted in the digital age. It has made access to information exponentially wider via social media but resulted not in the democratisation of news, but in 'information disorder'² that has been implicated in crises from fomenting genocide³ to insurrections⁴ to deaths by suicide.⁵

This landscape is no longer new: but policy is still slow to catch up. Our findings from this research suggests that the relationship between information producers and information consumers is broken. It is incentivised, shaped and enabled by a digital adtech ecosystem which prioritises engagement above all else, so that our public spaces for democratic discourse are not democratically functional. News and democratically important information is having to compete for audiences in spaces not built for it. And there is a proliferation of harmful discourse which cannot legitimately or practically be dealt with through current systems of content moderation.

In part one, we set out the findings from a large-scale data analysis of content posted on social media platforms (YouTube, Instagram and Facebook) by influential news organisations in the US and UK, and how audiences responded. In three case studies - stories surrounding trans influencer Dylan Mulvaney and her partnership with Bud Light, the controversies and charges against Hunter Biden (son of Joe Biden), and the fallout after the revelation from the TV presenter Phillip Schofield of an affair with a younger male colleague. Our key findings from this analysis are as follows:

Audiences are key drivers of digital discourse, but this operates within a social media engagement paradigm which does not lend itself to meaningful participation

- Amidst a focus on controversy and scandal, news and social media become the subject of the story themselves - from what media are reporting to what they are believed to not be reporting.
- Influencers, celebrities and politicians prolong controversies, taking their steer from online discourse and feeding further news coverage.
- Media is playing an important role in amplifying counterspeech. But this can lead to a backlash of further harmful discourse from audiences, such as articles platforming the calling out of transphobia resulting in further transphobia in response.

2 <https://firstdraftnews.org/long-form-article/understanding-information-disorder/>

3 <https://www.amnesty.org/en/documents/ASA16/5933/2022/en/>

4 <https://www.washingtonpost.com/technology/2021/10/22/jan-6-capitol-riot-facebook/>

5 <https://www.theguardian.com/technology/2022/sep/30/how-molly-russell-fell-into-a-vortex-of-despair-on-social-media>

Democratic discussion is intertwined with division, hate, and false or misleading narratives

- Political debate and democratic expression ends up intertwined with divisive narratives in audience discussions of news.
- True, misleading and weaponised information are all used to promote harmful narratives, from gendered disinformation to conspiracism.
- Audiences are not bound by news media reporting standards and so news fuels speculation, including making pejorative attacks and unsubstantiated allegations against individuals.

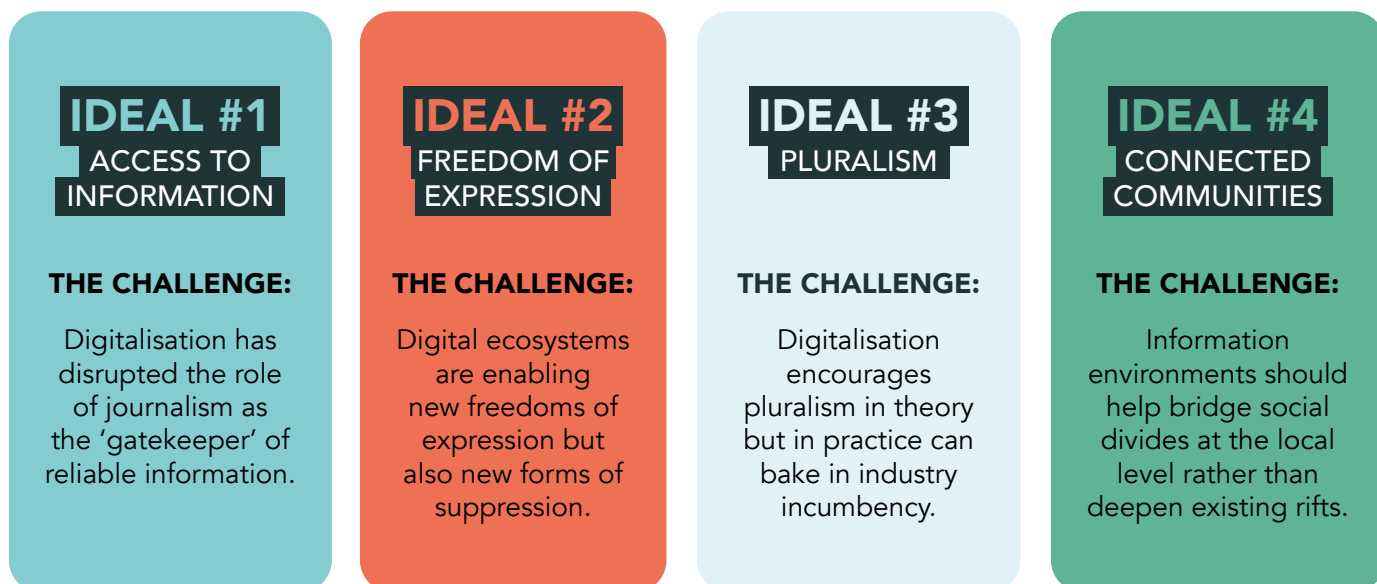
Information chaos persists, and people are not satisfied with the quality of democratic discourse or information production

- Distrust in information and information institutions is endemic, and spans from legitimate democratic critique to outright conspiracism.
- Media organisations and personalities being at the heart of the story becomes a wedge for conspiratorial narratives to gain a toehold, such as the media covering up government conspiracies or being involved in criminal enterprises.
- The distinction between news and opinion is blurry, as public figures' commentary frames factual reporting.

In part two, we set out four ideals of a healthy information environment and through expert interviews with journalists, academics, media and policy professionals, we explore how digitalisation and journalism are interacting to challenge these ideals. We set out the structural challenges that policy needs to address.

FIGURE 1

OUR FOUR IDEALS OF A HEALTHY INFORMATION ENVIRONMENT



In part three, we set out our recommendations of a pathway for change.

Building on our research insights, our recommendations are designed to move away from the current engagement-based paradigm which values clicks and attention regardless of the impact of the content that is being created and consumed. We need a new shared understanding of the different values that different kinds of content bring - not judged solely by their commercial worth, but by the common good they contribute to society and democracies.

We make five core recommendations designed to shift towards a new public interest news paradigm. First, we need to define what public interest news is. Attempts to do this have failed because of questions about who has legitimacy to define this. We propose a democratic solution to this:

RECOMMENDATION ONE:

The government should fund an independent People’s Commission on Public Interest News to involve the public in a deliberative way to devise a definition of public interest news

With this definition we can start to properly reward public interest journalism through public funding, new deals with social media companies to pay for and properly promote public interest news, and to start to build these values into the coming generative AI revolution that will again disrupt the information ecosystem.

RECOMMENDATION TWO:

Implement the Cairncross Review by providing public funding for public interest news

RECOMMENDATION THREE:

The development of digital competition legislation, including how news bargaining codes will work, should set out how the value of public interest news can be recognised by social media platforms, and such content appropriately promoted

RECOMMENDATION FOUR:

Media organisations should develop a standards code for companies producing generative AI tools to use their content

RECOMMENDATION FIVE:

The government should include a news bargaining framework for AI platforms in the forthcoming Digital Markets Bill to ensure that where news is being used in LLMs, news organisations are fairly compensated

We make three supplementary recommendations designed to address other aspects of the digital ecosystem.

RECOMMENDATION SIX:

The government should fund a major digital literacy programme across educational institutions but also pilot ways to educate other online audiences

RECOMMENDATION SEVEN:

The UK should follow the EU’s example and legislate to secure data access from major online platforms for public interest research

RECOMMENDATION EIGHT:

Ofcom’s advisory committee on disinformation and misinformation should take a wide view of information harms

We conclude by making the case we need to focus on how to rebuild the relationship between audiences, news producers and platforms. We need a new shared understanding of public interest news, to help challenge the engagement-based paradigm and shift towards a public interest-based paradigm that promotes social good and democratic discourse.

Finally, in the Annex, we set out the technical methodology for the large-scale data analysis undertaken for this project, including an account of methodological investigations carried out as part of this research.

GLOSSARY

Information environments: The spaces in which citizens access, consume, produce and interact with information. In this report, these are all digital.

Post: An original post on a social media platform by a news organisation, with text, either linking to an article posted on their own website, or accompanying a video hosted on the platform itself.

Reply: A comment on a social media platform in response to an article, either commented directly beneath the original post as a reply/comment, or text of a post which includes the link to the original post.


Harmful but legal speech: Speech which is likely to cause harm to an individual or group, either through causing psychological distress, by encouraging harmful behaviours, or being hateful speech which 'attacks or uses pejorative or discriminatory language with reference to a person or a group on the basis of who they are, in other words, based on their religion, ethnicity, nationality, race, colour, descent, gender or other identity factor,' but does not meet the threshold of illegality.⁶

Counterspeech: Speech which calls out or otherwise counters harmful speech, for instance by elevating the voices of those targeted.

News organisations: Organisations which produce and disseminate news media.⁷

CONTENT WARNING

This report includes discussion and example quotes relating to discrimination, hate and online abuse, as well as discussion of news stories which include potentially distressing themes including transphobia, homophobia, sexual abuse and death by suicide. Whilst we have kept this to a minimum as far as possible, we have maintained some examples to provide those working towards improving information environments with a strong and practical understanding of what we mean by harmful speech.

We have placed a  at the top, right hand corner of the relevant pages to indicate where this content appears.

Where quotes have been included from individual users in social media, these quotes have been bowdlerised in order to protect peoples' privacy. This means they have been copyedited to preserve the meaning but to prevent re-identification of the original user through search.

⁶ <https://www.un.org/en/hate-speech/understanding-hate-speech/what-is-hate-speech>

⁷ https://www.law.cornell.edu/definitions/uscode.php?width=840&height=800&iframe=true&def_id=2-USC-684343279-901562779&term_occur=1&term_src=title:2:chapter:26:section:1602

INTRODUCTION

Digital harms, from online violence to disinformation and misinformation, have been the centre of digital policy discussions for many years now. The drivers of online harms range from deep-rooted societal divisions to social media platform design for personal or political gain. As such, when we are considering how to respond to information harms we need to take a holistic view of the information landscape in which these are occurring.

This project aims to explore how news media interacts with social media, what the underlying drivers and impacts of this are, and what this means for how these interactions could be improved.

BACKGROUND

Our relationship with information has been uprooted. The digital age has made access to and distribution of information exponentially wider and easier by democratising sources of information, opening up new spaces for grassroots and citizen journalism, elevating the voices of historically excluded groups, and inculcating new forms of information production and consumption.

But at the same time, we are living in an age of 'information disorder'.⁸ Online information environments - particularly social media services - have been implicated in crises from fomenting genocide⁹ to insurrections¹⁰ to deaths by suicide.¹¹ The 'marketplace of ideas' that companies providing these services promise to support has been tested and sorely found wanting. And the explosion of development of generative AI technologies threaten a new disruption - for good or ill - to our information environments.

Alongside these seismic challenges the digital environment has brought to the consumption of journalism, it has also brought a profound disruption to the business models that underpin journalism. High quality journalism costs money and media companies are battling for financial sustainability. Yet social media algorithms don't necessarily reward public interest journalism, creating incentives that further exacerbate the problems.

People are increasingly finding their news through the medium of social media platforms, either as a space for consumption of and interaction with news content directly, or as a space to discover routes to news content. Though this gives news organisations new routes to engage diverse audiences, it means doing so in an environment designed by social media platforms for engagement, not meaningful news consumption. And the shrinking of other revenue avenues means that news organisations are increasingly dependent on these digital spaces.

Experts interviewed for this project highlighted how digitalisation has benefited some outlets, with organisations such as the New York Times combining commentator and opinion journalists with social media followings with a large subscription base. Smaller, local outlets, such as the Bristol Cable, have also built a successful financial model out of user support. But these models are extremely difficult to scale and standardise across all publications and rely on user subscription, which itself can pose a barrier for citizens to access news.

However, although social media platforms have disrupted traditional media and information dissemination, when discussing 'platforms', we often miss profound differences between the various social media sites, search engines, news aggregators, and messaging apps. Jonathan Heawood, Executive Director of the Public

8 <https://firstdraftnews.org/long-form-article/understanding-information-disorder/>

9 <https://www.amnesty.org/en/documents/ASA16/5933/2022/en/>

10 <https://www.washingtonpost.com/technology/2021/10/22/jan-6-capitol-riot-facebook/>

11 <https://www.theguardian.com/technology/2022/sep/30/how-molly-russell-fell-into-a-vortex-of-despair-on-social-media>

Interest News Foundation, made a distinction between the incentives for social media, such as Facebook, in news dissemination, compared to search engines such as Google Search, or news aggregators, such as Google News, NewsNow and Apple News. The incentive for search engines is to promote trusted news sources which result in positive user experiences. In comparison, social media tends to favour organisations which publish high volumes of content, focused on gossip and entertainment rather than public interest news.

For news organisations, this presents a dilemma: while they want to make sure journalism is available to wider audiences, higher-quality content is not favoured, we heard. Forward-thinking news organisations have always seen social media as an opportunity to reach new audiences. However, as some outlets have chased social media engagement, becoming reliant on traffic, they have found these sites to be fickle friends. There is a danger in becoming too reliant on traffic from social media, or even direct funding through various charitable initiatives,¹² as changes in the policy and design choices of these businesses can have a huge impact on the revenue of news outlets, and on freedom of expression more broadly.¹³ The recent closure of several prominent online news outlets, including BuzzFeed, Vice News and gal-dem, are a warning sign that high quality journalism, even when designed in and for the social media age, is not always sustainable.¹⁴

And in turn, this has left policymakers in a dilemma.

In the UK, the Cairncross Review¹⁵ was published four years ago and recommended new support for high quality independent journalism, with calls for online platforms to have greater responsibilities and oversight in their relationship with news and greater government support for public interest news. But the challenges have persisted and the policy conversation has moved from online safety regulation to digital competition regulation, data protection and AI regulation, with governments around the world now trying to wrestle back some control over the technologies. Too often, information and support for democratic discourse comes last on the list of priorities - safety and innovation are the buzzwords of tech policy, with difficult and contested issues such as how to design and promote healthy information environments left at the bottom of the barrel.

We need healthy information environments. That means free and wide access to high quality public interest news, support for independent journalism and empowered and digitally literate citizens who can participate and contribute in online spaces. It means safe, private online spaces, where communities can set the terms of engagement and diverse voices are given room to flourish.

But what is 'good' information, and who gets to decide? And how can we pay for it?

There are many tensions in this debate. Independent journalism plays a crucial role in the production, verification and curation of information and high quality journalism, accessible to all, and is the cornerstone of a democracy. But that role affords news producers huge power and can result in underrepresented voices being shut out.

The digital revolution in how people consume news promised to democratise information and amplify marginalised voices, but the spaces which have seen the rise of influencers and citizen journalists have also given greater platforms to extremists.

In an information crisis, promoting authoritative information, such as verified health information, can be a crucial lifeline for citizens. But entrenching top-down definitions of 'true' information can also reinforce conspiratorial ideas and overlook emerging knowledge.

Without legitimate levers to influence or compel corporations' decisions about content, engagement-based algorithms designed by platforms outside of any democratic scrutiny spiral hatred out of control. At the same time, government regulation is easily weaponised against citizens' interests as a means to block information access or content.¹⁶

Speech which degrades, dehumanises or threatens others is not conducive to the pursuit of freedoms, rights and equality. The limits of free speech need to be constantly tested, challenged and iterated as emerging harms and threats change. We have to ensure that the status quo of 'acceptability' is not used by those in power to simply impose their own limits of speech on citizens without accountability or recourse, or to shut out diverse voices from being heard

12 <https://www.openmarketsinstitute.org/publications/statements-on-deals-between-facebook-and-dominant-news-publishers>

13 https://demos.co.uk/wp-content/uploads/2022/05/Exemptions_Exceptions_and_Exclusions__OSB__vF.pdf

14 <https://www.vanityfair.com/news/2023/04/ben-smith-buzzfeed-news-traffic>

15 <https://www.gov.uk/government/publications/the-cairncross-review-a-sustainable-future-for-journalism>

16 <https://www.article19.org/resources/uk-online-safety-bill-serious-threat-to-human-rights-online/>

This is the minefield that has prevented meaningful change to improve our information environments. We need a way through this that accepts that although nothing is perfect, we can do better.

This report attempts to untangle these knots in our information environments for the better. We aim to describe what democratic discourse currently looks like, a vision for a better system and a pathway for change.

PART ONE

HOW NEWS MEDIA, SOCIAL MEDIA AND CITIZENS INTERACT TO DRIVE ONLINE HARMS

Social media spaces have become notorious for enabling, incentivising and amplifying a diverse array of online harms - from violence to abuse to disinformation. But social media platforms are not static spaces. They are dynamic information environments which bring together information producers and information consumers. Information producers, from traditional news media to influencers and content creators, compete for attention and engagement from digital audiences. Each can contribute to the health of the information environment, for better or for worse.

We wanted to investigate how news organisations were speaking about important issues in their online content, and how audiences were responding.

METHODS SUMMARY

In this report, we build on a large-scale analysis of social media data, covering a period from March to July 2023, on Facebook, Instagram and YouTube. We used a combination of natural language processing techniques in partnership with the University of Sussex to analyse the data, along with close qualitative analysis (a fuller account of the methods are included in the Appendix).

We collected posts made by 20 influential news organisations in the US and UK, from across the political spectrum over the course of this timeframe (see table below), then deployed topic modelling - an unsupervised machine learning technique - to group narratives and stories into digestible topics which could then further be analysed. We identified a hugely diverse range of subjects covered by news articles and videos - from political news about the workings of government, conflicts and crises around the world to celebrity dramas and lifestyle advice.

TABLE 1

SELECTED INFLUENTIAL NEWS ORGANISATIONS FOR COLLECTION OF SOCIAL MEDIA POSTS

UK NEWS ORGANISATIONS		US NEWS ORGANISATIONS	
• BBC News	• ITV	• ABC	• NBC
• Daily Mail	• The Mirror	• BuzzFeed	• The New York Times
• The Financial Times	• Sky News	• CNN	• People
• The Guardian	• The Sun	• Forbes	• The Washington Post
• The Independent	• The Telegraph	• Fox News	• The Wall Street Journal

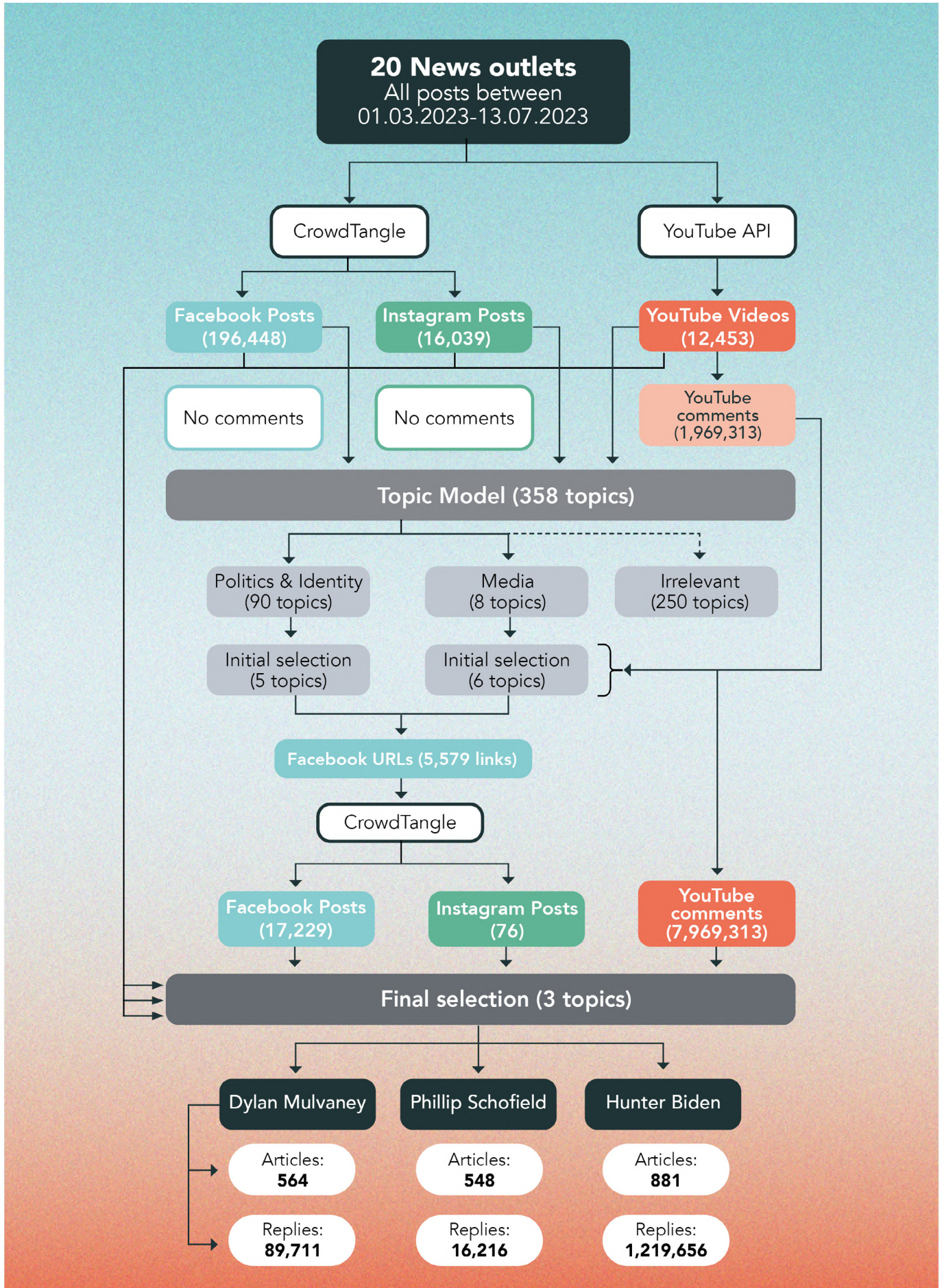
In line with our aim to understand how online harms are occurring in democratic discourse, we focused specifically on news organisation posts which covered political, media and topics relating to identity-based debates (such as about gender or race). We then collected replies to these posts from their digital audience - both replies made in direct response to a news organisation's post, and through new posts which linked back to the news story.

From this narrower selection of topics, we selected three highly engaged-with case studies of news stories across the US and UK markets to allow for a close in-depth analysis which took account of the individual contours of how different kinds of online harms manifest in different debates. These were different controversies surrounding public figures, including the son of US President Joe Biden, Hunter Biden; a British TV presenter, Phillip Schofield; and a TikTok influencer, Dylan Mulvaney.

In each case, we had a collection of replies which were used to produce new topic models representing the different types of narratives appearing as a response to the news stories. We developed qualitative coding frameworks through an iterative process, grouping the responses to the news stories into overarching themes. This allowed us to investigate what narratives were being shared about the story by news media organisations, what narratives we saw shared in response by audiences, and where these intersected in potentially harmful ways. We then drew on these insights, along with insights gathered through a series of expert interviews (see Part Two) to identify where changes could be made to how news organisations and digital audiences operate and interact to build a more positive democratic dialogue (see Part Three).

A flow chart showing the pipeline of data analysis which led us to these case studies is shown below.

FIGURE 1
DATA ANALYSIS PIPELINE



LIMITATIONS

A significant limitation we faced in this study was that of data access. During the time of the study, X (formerly Twitter) changed the terms of their data access, meaning that although it is a significant site of news interaction, it had to be excluded from our analysis. Similarly, although platforms such as TikTok are increasingly where consumers are accessing and interacting with news, data access limitations meant that these could not be collected.¹⁷ This is a significant limitation to the whole research community's ability to understand these public spaces, and we return to this issue in our recommendations.

Restrictions on the content discovery tool CrowdTangle, used to access Facebook and Instagram, also meant that we were not able to collect comments directly from Facebook and Instagram posts, and instead had to develop a proxy measure of collecting public posts which included a link to the original article, meaning not all reactions to these articles were included in analysis. Such link sharing was usually a less common way to react to an article than commenting on it; likely due to the fact that there is greater friction and barriers to actively resharing a link to an article than simply commenting, the limited link-sharing functionality on Instagram, as well as that our collection was only from public pages so shares on personal pages or in private messages would not have been collected.

Our analysis was text-based, which means that in the case of videos (on YouTube and Facebook), we relied on the text accompanying the video to describe the content of the video. In order to ensure uniformity between platforms, the methods treated YouTube video titles as 'headlines' and compared those to the titles of attached articles within the Facebook and Instagram data.

Our analysis was also limited to social media discourse, as this was our primary area of interest. However it is worth reflecting on whose voices may not have been represented in this, for example, audiences who commented on news stories in private messages, on other platforms, in person, or not at all, were not represented in this data.

In this section we set out our case studies, and what insights they offer into how news media, social media and citizens interact to drive online harms.

17 <https://www.reuters.com/business/media-telecom/fewer-people-trust-traditional-media-more-turn-tiktok-news-report-says-2023-06-13/>

HUNTER BIDEN

A CASE STUDY ON INSTITUTIONAL CORRUPTION, INJUSTICE AND CONSPIRACY

KEY FINDINGS:

1. Distrust in information and information institutions is endemic, with commenters vocally expressing distrust not just towards specific actors, but towards fundamental institutions and structures of governance.
2. Distrust manifests in legitimate democratic critique right through to outright conspiracism.
3. The distinction between news and opinion is blurry, as public figures' commentary frames factual reporting.

THE STORY

Hunter Biden, the son of President Joe Biden, has been the subject of numerous controversies and conspiracy theories.¹⁸ There have been legal proceedings against him on tax and gun charges, allegations that Joe Biden is involved in corrupt dealings related to his son's business contacts in Ukraine and China, as well as scandals and speculation about his personal life.

Emails from Hunter Biden's personal laptop relating to his role serving on the board of Burisma, a Ukrainian gas company, were leaked shortly before the 2020 US Presidential election in an article by the New York Post.¹⁹ Some argued that this was Russian disinformation, and Twitter and other social media sites blocked links to the story.²⁰ Former CIA/National security officials signed a letter stating no evidence of Russian disinformation, but airing their suspicions that this was part of a Russian information operation.²¹ However, the emails turned out to be genuine (though it remained unclear how they got into the hands of Biden's political adversaries) and these became relevant to the charges brought against Hunter Biden.

Discussion of these events occurred against a background of deep suspicion and mistrust in the veracity of information from all sides. As part of House Oversight Chairman James Comer's investigation into Biden family, he claimed that 'America witnessed a coordinated campaign by social media companies, mainstream news and the intelligence communities to suppress and de-legitimized the existence of Hunter Biden's laptop and its contents,' and that Twitter 'worked hand-in-hand with the FBI to monitor the protected speech of Americans, receiving millions of dollars to do so.'²²

18 <https://www.bbc.co.uk/news/world-us-canada-55805698>

19 <https://nypost.com/2020/10/14/email-reveals-how-hunter-biden-introduced-ukrainian-biz-man-to-dad/>

20 https://apnews.com/article/business-media-social-media-censorship-ec529ef85c1e72cefe0ae9450e118b9c?utm_source=apnews&utm_medium=relatedcontentmodule

21 <https://www.politico.com/news/2020/10/19/hunter-biden-story-russian-disinfo-430276>

22 <https://edition.cnn.com/2023/02/08/politics/twitter-hearing-house-oversight/index.html>

HOW IT SPREAD ONLINE: ENGAGEMENT AND KEY THEMES OF DISCUSSION

The initial posts collected about this story covered a wide range of subjects, including discussion of Joe Biden’s presidency and administration, a wide range of Democratic politicians, Hunter Biden, Donald Trump, the probe into Trump’s alleged collusion with Russia, and Trump’s indictment. Due to the high volume of data for this topic, posts were filtered using a combination of exact phrase matches (e.g. “Hunter Biden”) and broader combinations of words (e.g. any post using both words “Biden” and “laptop”), and then further qualitatively annotated to narrow the dataset down to the most relevant posts - that which focused on controversies and allegations around Hunter Biden and the Biden family specifically (excluding, for instance, political criticism of Joe Biden’s policies but including conspiracy theories about his family).

Unsurprisingly, the vast majority of this coverage was by US outlets, with 15 posts in total from UK media outlets. On YouTube, CNN was the most engaged with outlet: receiving the most replies on average per video (an average of 4558 per video vs 3138 per video for the second most, Fox News). On Facebook however, Fox News was the most engaged with outlet, receiving the most likes on average per post (1295 likes per post vs 706 for the second most, CNN), and most shares on average per post (an average of 324 shares vs 209 for the second most, Daily Mail).

The majority - around two-thirds - of these posts from media outlets were on YouTube, with the remaining third posted on Facebook.

Between March and mid-June 2023, the majority of the coverage was from right-leaning outlets, however there was a significant spike in coverage from centrist outlets in late June-July. The story was much less frequently covered by left-leaning outlets at all until this same period of late June. This period in June was when Hunter Biden was charged with tax and gun crimes.²³

This speaks to the political polarisation which surrounded this story: within the dataset we saw significant discussion of the claim that the story ‘had not been covered’ by powerful left-leaning media organisations, feeding the idea of a cover-up. Headlines mirrored this division, with left leaning outlets focusing on the alleged lack of relevance of the story, and right leaning outlets focusing on the allegations of a cover-up.²⁴

The main narratives present in the posts around the Hunter Biden story were as follows:

TABLE 2
MAIN NARRATIVES OF DISCUSSION RELATED TO HUNTER BIDEN WITHIN OUR DATASET OF POSTS FROM NEWS ORGANISATIONS

NARRATIVE	DESCRIPTION	EXAMPLE HEADLINES	NUMBER OF POSTS
Other narratives	<p>Included general discussion of the various figures involved in the legal case against Hunter Biden, spanning lawyers, minor political figures within both US political parties, and the Biden family.</p> <p>Posts were often long and spanned several very specific aspects of the story, hence our decision to focus on specific, easily identifiable areas of discussion for deeper analysis.</p>	<p>‘IRS whistleblowers to testify on alleged meddling in Hunter Biden case IRS whistleblowers to testify about alleged meddling in Hunter Biden case at House hearing</p> <p>The public hearing is expected to focus on testimony from former IRS criminal investigator Gary Shapley and a second unnamed IRS criminal investigator, who, the panel says, possess ‘critical information’ related to their probe into the Biden family.’</p>	769

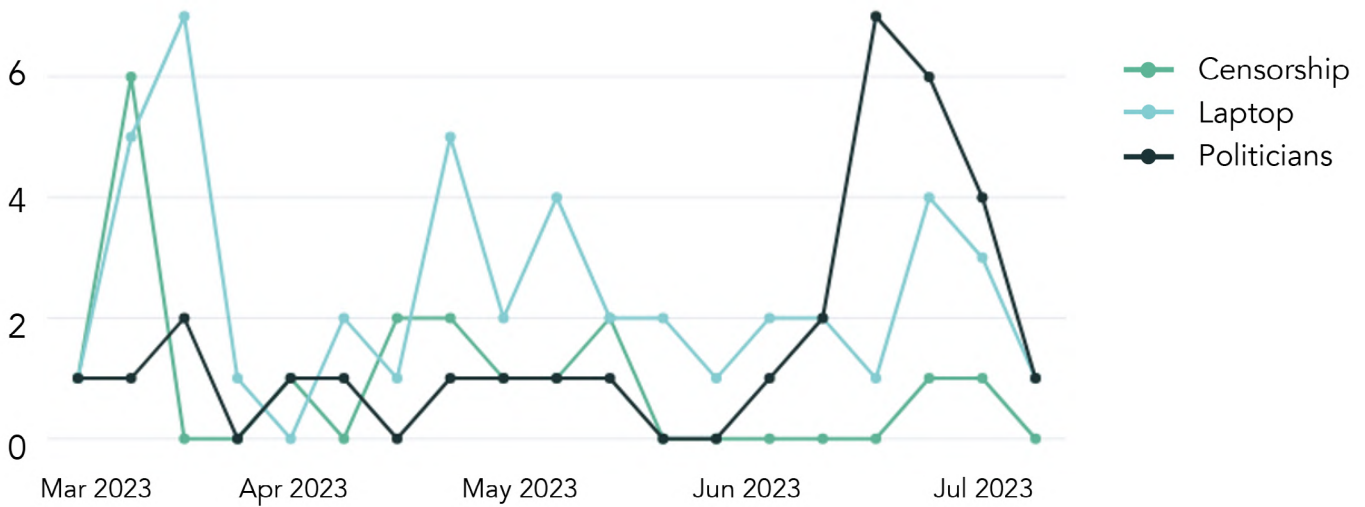
²³ <https://www.justice.gov/usao-de/pr/tax-and-firearm-charges-filed-against-robert-hunter-biden>

²⁴ To determine political leaning of outlets, we drew on previous research from the Pew Research Center <https://www.pewresearch.org/journalism/2021/04/28/biden-administration-100-days-appendix-a-grouping-outlets-by-audience-ideology-and-grouping-survey-respondents-by-media-diet/>

NARRATIVE	DESCRIPTION	EXAMPLE HEADLINES	NUMBER OF POSTS
Laptop	<p>Discussion specifically related to Hunter Biden's laptop and the material which was leaked from it (emails, messages, photos, videos), and the subsequent use of the material in legal proceedings.</p> <p>Included posts on the controversy over whether the laptop story itself was true or relevant, and whether there was a conspiracy to dismiss it as disinformation.</p>	<p><i>'Remember The Laptop From Hell?': Trump Puts Hunter Biden In Crosshairs In Post-Indictment Speech'</i></p>	50
Politicians	<p>Included significant coverage of Trump's remarks about Biden and the Biden family, as well as other lawmakers' comments such as Ted Cruz. Common themes included not merely criticism of Biden but of the entire US justice system, cover-ups, and protection of the Bidens by institutions such as the DOJ or FBI.</p>	<p><i>'DOUBLE STANDARD: Former President Trump tells Fox News that President Biden and his family are 'being protected' by the 'corrupt' and 'one-sided' justice system in the United States.'</i></p>	35
Censorship	<p>Discussion related to the removal of the New York Post story which first leaked the material from Hunter Biden's laptop from various social media sites, following a public statement written by several former intelligence officials which stated that the materials may be part of a Russian disinformation campaign, ahead of the 2020 US election.</p> <p>Discussed 'suppression' of information and alleged coordination to get signatories onto the letter discrediting the laptop story including the Secretary of State, the CIA and the Biden for President campaign. It also included analysis of these allegations and posts claiming that the allegations of censorship, suppression or coordination were not convincing.</p>	<p><i>'Dan Goldman Grills Twitter Files Journalists, Matt Gaetz Blasts Suppression Of Hunter Biden Story</i></p> <p><i>At today's House Weaponization of the Federal Government Committee hearing, Rep. Dan Goldman (D-NY) questioned Twitter [...]</i></p>	21

FIGURE 2

VOLUME OF ARTICLES OVER TIME RELATED TO 3 KEY NARRATIVES (CENSORSHIP, LAPTOP AND POLITICIANS) WITHIN OUR HUNTER BIDEN DATASET



This was by far the most highly engaged-with story we investigated, with over 1.2 million replies, predominantly from YouTube: likely due to a combination of the fact that this was predominantly covered by Fox, which has a strong YouTube presence, and our collection restrictions which meant that Facebook comments could not be collected directly.

The narratives which came out in the replies had a common theme: distrust and corruption.

TABLE 3

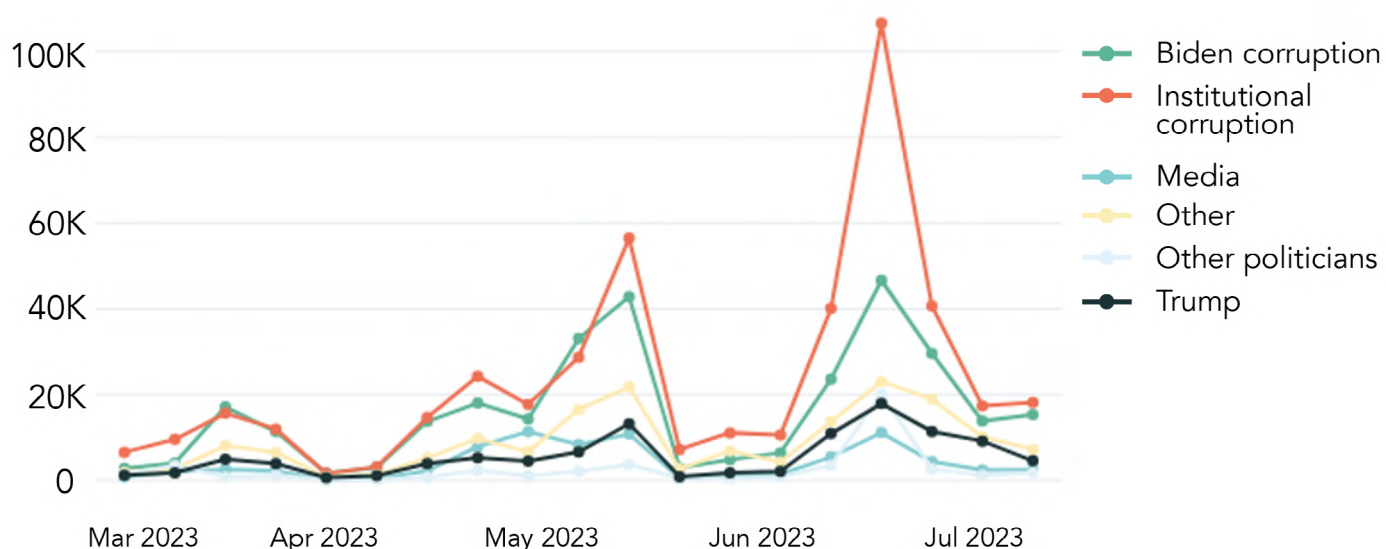
MAIN NARRATIVES OF DISCUSSION WITHIN OUR DATASET OF USER REPLIES TO NEWS ORGANISATIONS' POSTS RELATED TO HUNTER BIDEN

THEME	DESCRIPTION	EXAMPLE COMMENTS (BOWDLERISED)	NUMBER OF COMMENTS
Biden corruption	Discussion of corruption specific to the Biden case, naming relevant actors (the Bidens, Hunter Biden, President Biden).	'Evidence is mounting up against Biden Crime Family. It is beyond all belief that there haven't been indictments, arrests or prosecutions!'	316230
Corruption of the justice system	Relating to the FBI, CIA, Department of Justice, and the overall criminal justice system - including legal proceedings, lawyers, and failure to prosecute criminals.	'For how long are the ATTORNEY GENERAL and FBI gonna be allowed to lie to Congress and the American people? Its time to imprison and impeach'	215652
Other	Miscellaneous, generally irrelevant or not useful for analysis.	'Today's Republican Party is a bunch of clowns 🤡🤡🤡'	175438

THEME	DESCRIPTION	EXAMPLE COMMENTS (BOWDLERISED)	NUMBER OF COMMENTS
General distrust in institutions	Relating to a more generalised sense that nothing will change, everyone is a liar, everything is rigged, without naming specific 'bad actors' or entities.	'I do not and will not hope because nothing ever happens. So why keep going on and on. Nothing ever happens... nothing'	155178
Trump	Discussion of Trump in relation to the Hunter Biden case - i.e. his comments on the situation - or general mentions of Trump - i.e. vote for Trump in 2024.	'President Trump's the only man on the planet who can save our country and every one knows it.'	111521
Corruption of democratic systems	Relating to political parties, politicians and elections.	'All of the politicians are involved, that is how it persists so long as it does.. It's called a uniparty for a reason.. Every thing else is just a circus for the benefit of us'	97101
Media	Discussion of Fox News, specific media figures such as Tucker Carlson.	'lol...Just keep believe the FOX NEWS propaganda... 🤪'	79885
Other politicians	Discussion of other prominent US political figures, as above.	'Wow... Senator J Hawley... Straight forward and excellent question Verybrilliant. Excellent job.'	54852
Conspiracy leaning	Relating to the more extreme fringes of institutional distrust, where individuals are spreading established conspiratorial narratives, including QAnon, climate change denial, and covid-19 denial.	'the wars a racket. climate change/global warmings a con. covid-19 is a con. just follow the money. that's all thats needed.'	13799

FIGURE 3

VOLUME OF COMMENTS OVER TIME WITHIN OUR HUNTER BIDEN DATASET, SHOWING ALL MAIN NARRATIVES



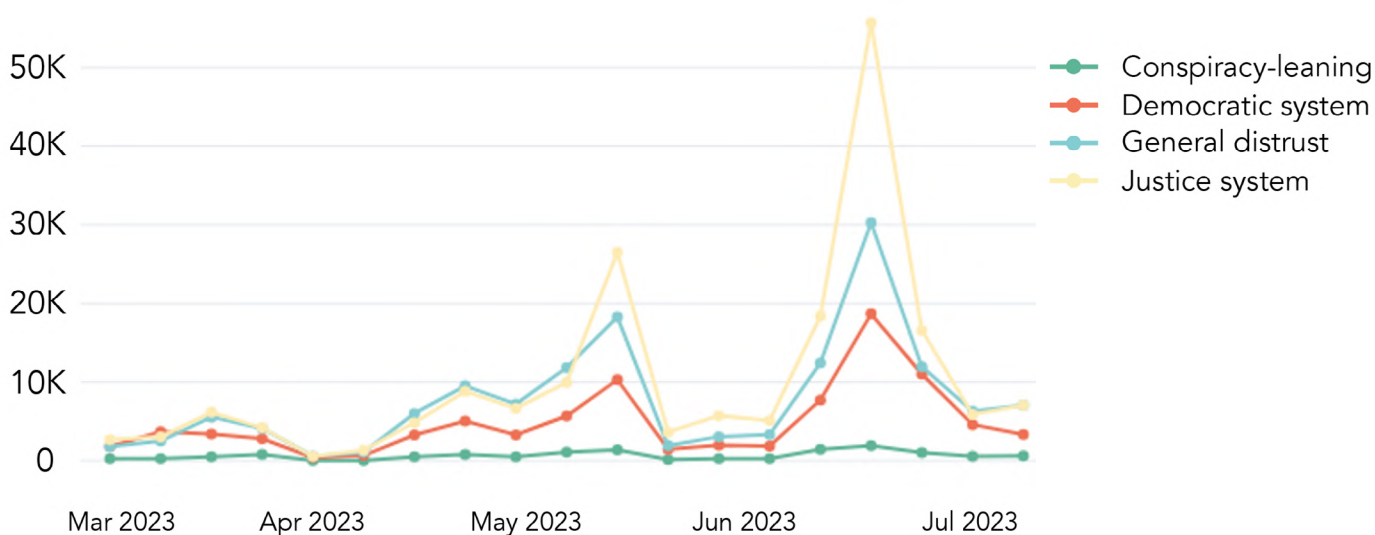
As shown in the graph above, the most prevalent narrative was Institutional Corruption, which covered discussion of systemic and institutional corruption. This ranged from a belief that officials and institutions are failing, through to being actively controlled by hostile actors in order to oppress citizens.

Comment: *'It isn't about Trump - it is all about the systemic corruption of all US institutions - the military, the law enforcement, the judicial and the civil - and all the time, at the heart of all the pyramids of corruption are just the same kind of people'*

In order to better illustrate the diverse range of issues covered by the theme of Institutional Corruption, we went through another stage of qualitative investigation to break this down into more granular categories: Corruption of the justice system, Corruption of democratic systems, General distrust in institutions and Conspiracy-leaning.

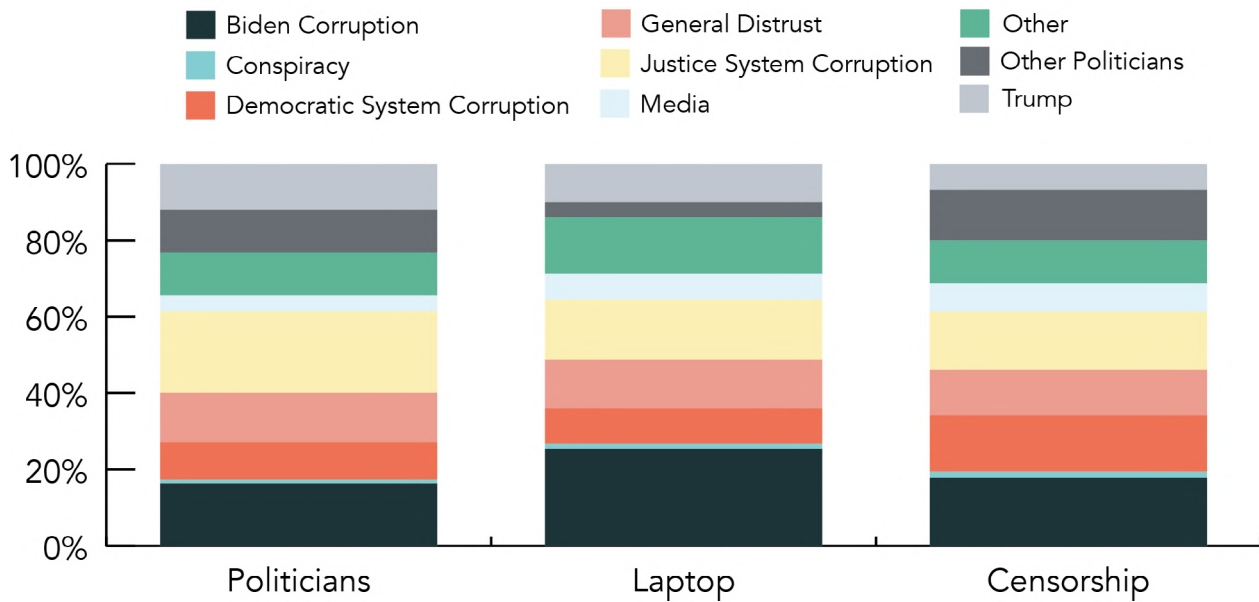
FIGURE 4

VOLUME OF COMMENTS OVER TIME WITHIN OUR HUNTER BIDEN DATASET, SHOWING THE BREAKDOWN OF THE NARRATIVE OF 'INSTITUTIONAL CORRUPTION' INTO SUB-NARRATIVES



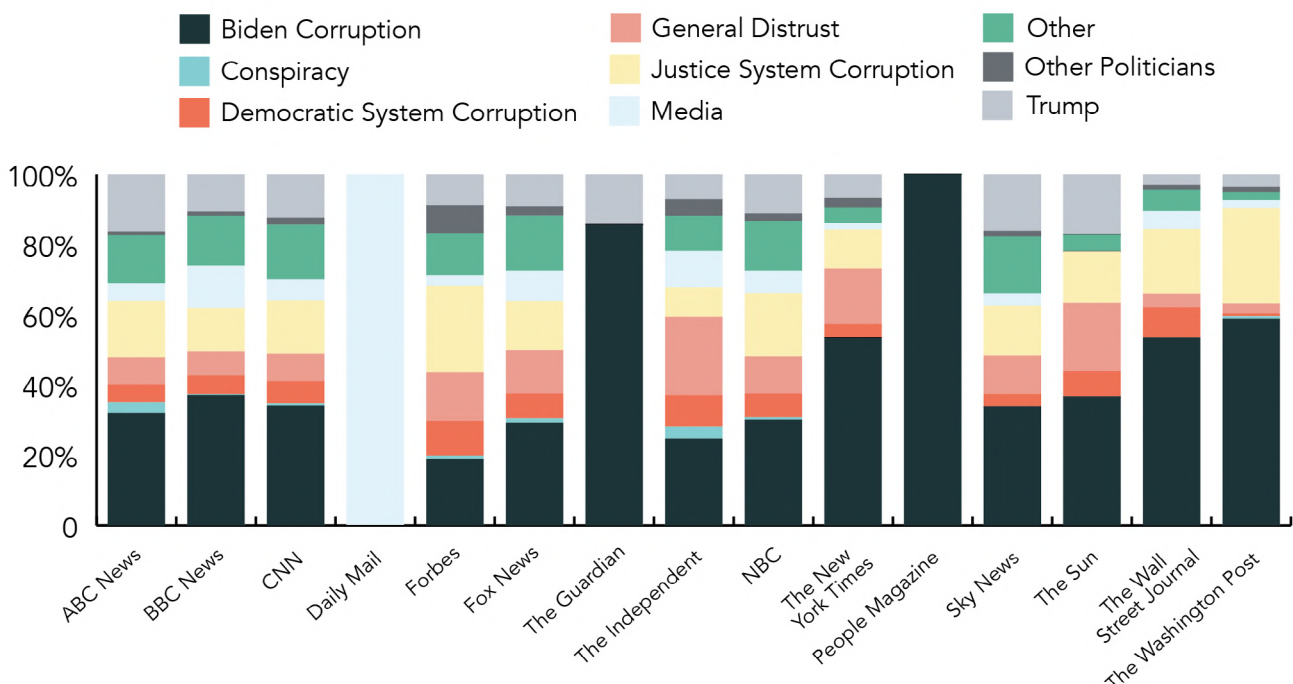
For each of the different narratives present in the posts from news organisations - Politicians, Laptop and Censorship - we found similar proportions of each discourse type within the corresponding replies, with corruption related to the Biden family and to the justice system most prevalent. This suggests that audiences are primarily expressing their own pre-existing views when interacting with news, rather than being particularly directed by the framing of news headlines.

FIGURE 5
 PERCENTAGE OF REPLIES TO POSTS FROM EACH OF OUR KEY 3 NARRATIVES (CENSORSHIP, POLITICIANS, LAPTOP) WHICH CONTAINED EACH TOPIC



The proportions of each discourse type within the corresponding replies to specific news outlets showed more variation, though this appears largely due to skewed distribution of articles from particular outlets. For example, there were very few posts made on this topic by the Daily Mail, People Magazine or the Guardian.

FIGURE 6
 PERCENTAGE OF REPLIES TO POSTS BY NEWS ORGANISATIONS WHICH CONTAINED EACH TOPIC



KEY FINDINGS

This discourse should, in theory, be the epitome of democratic and public interest discourse: identifying corruption, criminality or systemic injustices occurring at the heart of government, exposing cover-ups and inconsistencies, and enabling the public to hold their elected officials to account for any misdeeds.

However, what we see is information chaos. Each side defends their own and attacks the other, and commenters respond by not trusting anything they see, be it media, government or otherwise. We see scepticism of Biden overlapping with scepticism of all government institutions and moving into full-blown conspiracies. This raises concerns that this paves the way for potential radicalisation as people who legitimately criticise the government find themselves in an information ecosystem where that criticism is put on a level playing field with wide-ranging and even extremist scepticism.

The distinction between news and opinion is blurry, as public figures' commentary frames factual reporting

We examined the nature of the coverage (based on the text of the social media post of the article) to qualitatively assess whether articles were primarily being presented as opinion articles. This includes those with headlines specifically labelled as opinion, analysis or commentary, as well as ones which quote media figures or feature videos of media hosts discussing a story, or news articles featuring direct reporting of events, or quotes from prominent figures such as politicians.

News

Headline: BREAKING NEWS: Hunter Biden Will Plead Guilty On Tax Charges—Avoid Gun Charge

Opinion

Headline: Republicans crying wolf over Hunter Biden have hurt their own cause

A key finding was that this distinction was not always simple to draw. Some outlets clearly labelled their analysis pieces as 'opinion' or 'analysis'; however, many did not, and combined editorialising comments with reporting events. A common occurrence was outlets using a quote as the headline or beginning of the headline, which was reporting someone else's commentary or opinion about the event in question, thus amplifying that position while still reporting.²⁵

Headline: "DISGUSTING": Comer to hold press briefing on Biden family business dealings'

We also reviewed, across both primarily news and primarily opinion posts, whether the language used was neutral or evaluative, meaning using emotive or evocative language drawing a moral judgement on the events or people discussed. This further blurred the distinction between news reporting and opinion, as there were news reports which used highly emotive language, and opinion reporting which was presented in neutral language.

The risk of this is that it is unclear to audiences what role different forms of content are playing and how they have been produced - where a quote, for instance, is an opinion of the author vs the opinion of another commentator, such as a politician, being reported, and what is evidence vs a subjective assessment of the situation.

Opinion, Evaluative

Headline: 'BEYOND INSULTING': Sean Hannity blasts Hunter Biden's plea deal, says AG Merrick Garland is a 'water boy for the sleazy, Biden family syndicate in Delaware.'

News, Evaluative

Headline: JUST IN: Trump Explodes On Joe And Hunter Biden Over Cocaine: 'Was Crooked Joe Biden On Cocaine...?'

²⁵ Where an issue was being discussed by a media figure, or someone's comments were being reported as part of a media show (example), it was classified as 'opinion', whereas when it was reporting events or reporting the fact of someone else's commentary outside of a media context, unless clearly editorialising in the headline, it was labelled as 'news'.



Emotive framings had higher levels of engagement, as did posts with headlines featuring major, well-known political figures: which tracks the highly polarising and high-profile nature of this particular topic. This also enables public figures to further exploit existing divisions by controlling the narratives which will be reported on through their commentary.

Distrust in information and information institutions is endemic

This dataset contained the full range of institutional mistrust: from legitimate questions and concerns about the operations of government, and real events which undermined people’s trust in the systems of justice and fairness, to outright conspiracism. Political divisions over which ‘side’ is telling the truth or spreading disinformation are amplified in news reporting and audience responses, leading to legitimate and fabricated allegations of wrongdoing being levelled at all players by citizens.

The themes which arose from the replies to posts about Hunter Biden showed huge levels of distrust in all directions, in particular right-wing distrust of government institutions such as the FBI, CIA, or DOJ, of the mainstream media, and of Democrats and the current government. Also present was criticism of Trump, the MAGA movement, Fox News, and the Republican party, from all political leanings.

Distrust spans from legitimate democratic critique to outright conspiracism

Replies focusing on institutional corruption expressed distrust of everything from the results of the 2020 elections, to the existence of the Covid-19 pandemic, to extreme fringes of conspiracy theories, including QAnon conspiracies and accusations of child sex trafficking or even child sacrifice by politicians. Discussion of corruption by the Biden family tended to be more narrowly related to the story, expressions of the allegations against Hunter Biden specifically, but also speculation about the involvement of President Biden. There was also criticism of mainstream media, including accusations that the Hunter Biden laptop story was censored.

Very few replies engaged in debunking conspiracy theories, or in defending the Bidens. Those which were more ‘pro-Biden’ tended to express that by pointing out hypocrisy, accusing Trump and his children of similar or worse corruption.

HEADLINE	COMMENT (BOWDLERISED)
<p>‘BREAKING NEWS: Ted Cruz Explodes On Top FBI Official Over Biden ‘Bribery Scheme’ Allegations</p> <p>At today’s Senate Judiciary Committee hearing, Sen. Chuck Grassley (R-IA) questioned Deputy FBI Director Paul Abbate about ...’</p>	<p>‘Another Swamp Creature...Abolish FBI’</p> <p>‘This guy’s a crook & thinks he’s above the law. Send his a55 to prison for treas0n against the country. Thank you Mr Cruz for kicking his butt and staying on top of the dirtbag. DirtyDurbin’s a snake in the grass & need to be charged. I am tired of all the lies and B S excuses those criminals make. But there is a pay day coming for all of the crooks.’</p>
<p>‘BREAKING NEWS: Trump Announces He’ll Appoint Special Prosecutor To Go After ‘Biden Crime Family’</p>	<p>‘Fox news is also complicit. Fox news is bringing violence to the most vulnerable among us. Civil rights as Black, Brown, women, lgbt or democrat citizens are trampled at Fox News. Hannity is ageing, Linsey has the vapours and Fox News has been a non-stop SNL skit since the indictment. I think old Jack Smith is going to be talking to Ingram, TucKKK er and Gutfeld. Fox is a national-security-threat to the US.’</p>



Although it was not hugely prominent, there was also evidence within the Hunter Biden data of more extreme conspiracy theory narratives.

Comment: *'It had been a conspiracy theory that the virus had been man made - and we know now, it's true! Mainstream media and big tech companies are working over time to ban and censor it. It was released deliberately to fuck up the American election and so China could get the protesters off of the streets of HK...'*

Comment: *'Most elites have involvement in trafficking organs and kids'*

Our hypothesis is that feeling unrepresented in mainstream media lends itself to conspiracism amongst audiences. Fox News has a gigantic market share in the US, with no mainstream conservative competitor,²⁶ and being primarily broadcast, unlike most of the other outlets examined, has enabled it to make a very successful move to YouTube, with 10.8 million subscribers at time of writing. This 'gives Fox incredible power'.²⁷ However, Fox News has not been untouched by the populist backlash against the 'MSM': after Fox News fired Tucker Carlson in the aftermath of a defamation lawsuit against Fox News by Dominion, the voting equipment company, for their repeated broadcast claims that the election was stolen (although it is not known if these two events were connected).^{28,29} Many of Fox's core base felt alienated by this firing. Court documents from the lawsuit showed that media hosts on Fox felt pressure to cave to their audiences' belief that the election was stolen.³⁰

This is a clear example of media being audience-led but not in a positive way or in the public interest. The challenge for media is how to compete with alternative media outlets which will cater to audiences' beliefs but can be further isolating and radicalising. There is a need to build a different kind of relationship with audiences.

26 Though some have argued that Newsmax has the potential to emerge as a significant rival for Fox News, i.e. <https://www.newyorker.com/news/annals-of-communications/what-dominion-has-to-prove-in-its-case-against-fox-news>

27 A quote from our interview with Professor Jeff Jarvis, Director of the Tow-Knight Center for Entrepreneurial Journalism at the City University of New York.

28 <https://www.theguardian.com/us-news/2023/apr/18/fox-dominion-settle-us-defamation-lawsuit>

29 <https://www.theguardian.com/media/2023/jun/24/fox-news-tucker-carlson-dispute>

30 <https://www.nytimes.com/2023/02/16/business/media/fox-dominion-lawsuit.html>



PHILLIP SCHOFIELD

A CASE STUDY ON THE MEDIA UNDER SCRUTINY

KEY FINDINGS:

- Amidst a focus on controversy and scandal, news and social media become the subject of the story themselves.
- Audiences are not bound by news media reporting standards and so news fuels speculation.
- Media organisations and personalities being at the heart of the story became a wedge for conspiratorial narratives to gain a toehold.

THE STORY

Phillip Schofield, the UK TV presenter of *This Morning*, was in the news this year (2023) over revelations about a relationship with a younger male colleague at work³¹ and his leaving the show.³² There was significant backlash leading to online attacks as well as other media personalities becoming involved in the debate on different sides, some defending and others critiquing Schofield and his employer and maker of *This Morning*, ITV.

HOW IT SPREAD ONLINE: ENGAGEMENT AND KEY THEMES OF DISCUSSION

The dataset of posts relating to the story was broad, including posts about *This Morning* more widely than just the affair, and including posts about Phillip Schofield's brother being convicted for child sexual abuse. As such, the dataset was filtered to include those stories which were most clearly relevant from the headline to the affair itself, as a source of particular controversy which generated great debate with a clear impact.

Amongst the stories in our dataset which had a UK focus, this was the most engaged-with story. Unlike the US stories, which also were covered by UK outlets, this story was covered exclusively in the UK, with the exception of 1 story by the Washington Post, and covered more frequently by left-leaning outlets, most significantly the Mirror. These were posted much more heavily on Facebook than on other platforms.

Coverage of the relationship, or of a story related to it, had several different framings, or lenses through which the story was approached. Although there were many overlaps between categories, below is a breakdown of the prominent themes and which frames posts primarily centred:

31 <https://www.independent.co.uk/arts-entertainment/tv/news/phillip-schofield-statement-this-morning-b2348007.html>

32 <https://www.theguardian.com/uk-news/2023/may/26/phillip-schofield-quits-itv-after-admitting-affair-with-this-morning-colleague>

**TABLE 4**

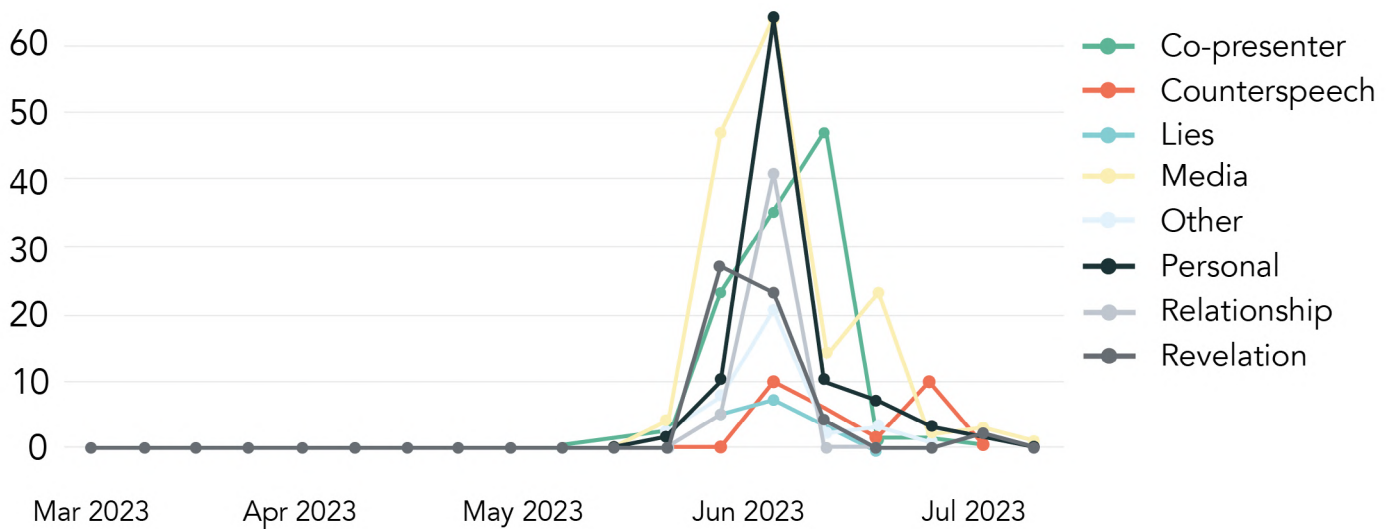
MAIN NARRATIVES OF DISCUSSION RELATED TO PHILIP SCHOFIELD WITHIN OUR DATASET OF POSTS FROM NEWS ORGANISATIONS

NARRATIVE	DESCRIPTION	EXAMPLE HEADLINES	NUMBER OF POSTS
Media	Focus on actions of media organisations, most prominently ITV	'ITV chief executive Dame Carolyn McCall asked to appear before MPs to face questions over the Phillip Schofield row'	160
Co-presenter	Centering Holly Willoughby (his co-presenter)	"Hurt' Holly Willoughby breaks silence on Phillip Schofield's affair and slams former pal'	117
Personal impact	Focusing on the personal mental health on Schofield and colleagues, career or family impact of the story	'Phillip Schofield compares himself to Caroline Flack as he says 'I've lost everything''	107
Revelation	Focusing on the 'revelations' of Schofield's statements or interviews	'Phillip Schofield's full statement as he reveals he lied about This Morning colleague affair'	57
Relationship	Focusing on details of facts about the relationship itself, including focus on age of the colleague and when they met	'Inside Phil Schofield's affair - 'playtime, bolt hole and sleepovers at wife's home'' 'This Morning employee 'was 15' year old when he met Phillip Schofield during school visit'	46
Other	Other more general stories about Schofield or the fallout	'Phillip Schofield lashes out at 'people with grudges' in first Instagram post since affair scandal'	38
Counterspeech	Amplifying or engaging in counterspeech against attacks, false allegations or homophobia	'Phillip Schofield slams affair 'homophobia' as he compares himself to Leo DiCaprio'	29
Lies	Focusing on admissions or allegations of lying	'Eamonn Holmes slams Phillip Schofield saying 'he's lied to everyone' and 'more will come out''	15



FIGURE 7

VOLUME OF ARTICLES OVER TIME WITHIN OUR PHILIP SCHOFIELD DATASET, SHOWING ALL MAIN NARRATIVES



Posts followed the same pattern as the other two case studies in that they covered both the initial events themselves, how those directly involved were responding or speaking about the case, and then how secondary figures, other commentators, politicians, celebrities and so on, had talked about the events. Media personalities featured strongly in the reporting, with posts focusing either on comments made by or the impact on his co-host Holly Willoughby, as well as other This Morning presenters, including Alison Hammond and Dermot O’Leary.

These post narratives broadly covered two categories: looking at the wider social implications or impact of the affair, and looking at the personal details of the relationship itself and the personal impact on those involved.

The personal details commonly shared included details of the relationship itself, which were sometimes described in terms which evoked a sense of ‘scandal’:

Headline: ‘Phillip Schofield selling plush flat ‘where he had secret trysts with his toyboy lover’’

Reportage also included discussions of the mental health impact on those involved most closely, including statements made officially, as well as sharing further details of the activities of those involved.

Headline: ‘I’m mentally utterly broken, if it hadn’t been for my daughters Ruby & Molly I wouldn’t be here, says Phillip Schofield’

The social impact themes commonly discussed were around workplace relationships and safeguarding, and what policies or processes should be in place to protect people, particularly in relationships involving power imbalances. This included both analysis and reporting content:

Headline: ‘Phillip Schofield: ITV boss called to answer questions by MPs on safeguarding after star’s affair’

Headline: ‘The lesson from the Phillip Schofield scandal? A moral grey area is not OK in any workplace’

Replies covered similar themes, but frequently focusing on perceived wrongdoing, either of individuals or the organisations involved.

**TABLE 5**

MAIN NARRATIVES OF DISCUSSION WITHIN OUR DATASET OF USER REPLIES TO NEWS ORGANISATIONS' POSTS RELATED TO PHILIP SCHOFIELD

THEME	DESCRIPTION	EXAMPLE COMMENTS (BOWDLERISED)	NUMBER OF COMMENTS
General	Responding to specific aspects of a story	'Fact that he's going round to do self-publicising interviewing, it is mad'	8029
Media	Criticism of the media response to the story or actions of media organisations such as ITV	'So the BBC tries to destroy Andrew Tate but goes easy on Scofield? Media are just puppets, and it's the govmt who pull their strings'	2270
Grooming	Discussing/making allegations/accusations of paedophilia, grooming, criminal behaviour and abuse	'One more Saville & Harris...'	1806
Schofield	Discussing Phillip Schofield and the story generally, across multiple aspects	'Schofield's just reading over 1,000,000s of the YouTube comments about this, he'll reply to you soon'	1477
Relationship	Discussion of whether the relationship was appropriate, particularly focusing on the age gap and power dynamics, with people both defending and criticising the legality or appropriateness of the relationship	'Not to be funny, if that were a man with a younger woman, it would all be okay'	1063
Lies	Focusing on admissions or accusations of lying	'Lying, lies + further lies, to make it seem okay'	764
LGBT	Discussion of Schofield's sexuality, including homophobia, defence of Schofield and general discussion of whether sexuality is relevant to the story and whether responses to the story are homophobic	'Trying and spinning this into 'shaming gay people' is crazy...'	464
Mental Health	Discussion of the impact of the story on Schofield's mental health following interviews about it. This included both support for him as well as further critique	'I do hope that he can get all the support for his mental health that is needed. I hope he can come back after this.'	343

FIGURE 8

VOLUME OF COMMENTS OVER TIME WITHIN OUR PHILIP SCHOFIELD DATASET, SHOWING ALL MAIN NARRATIVES

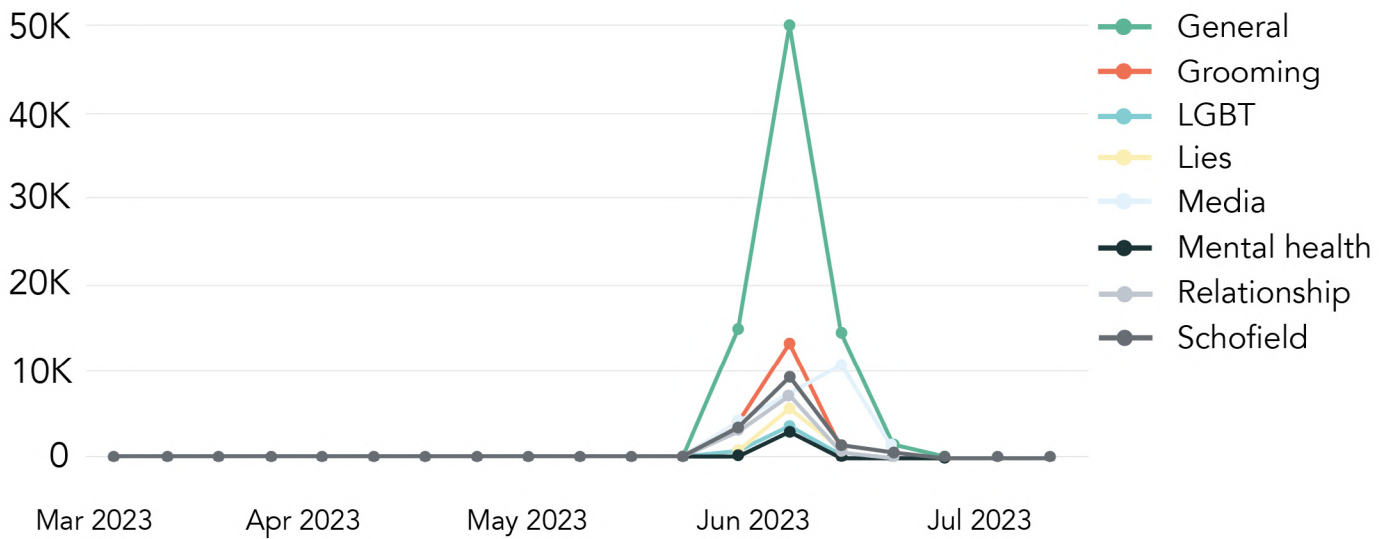
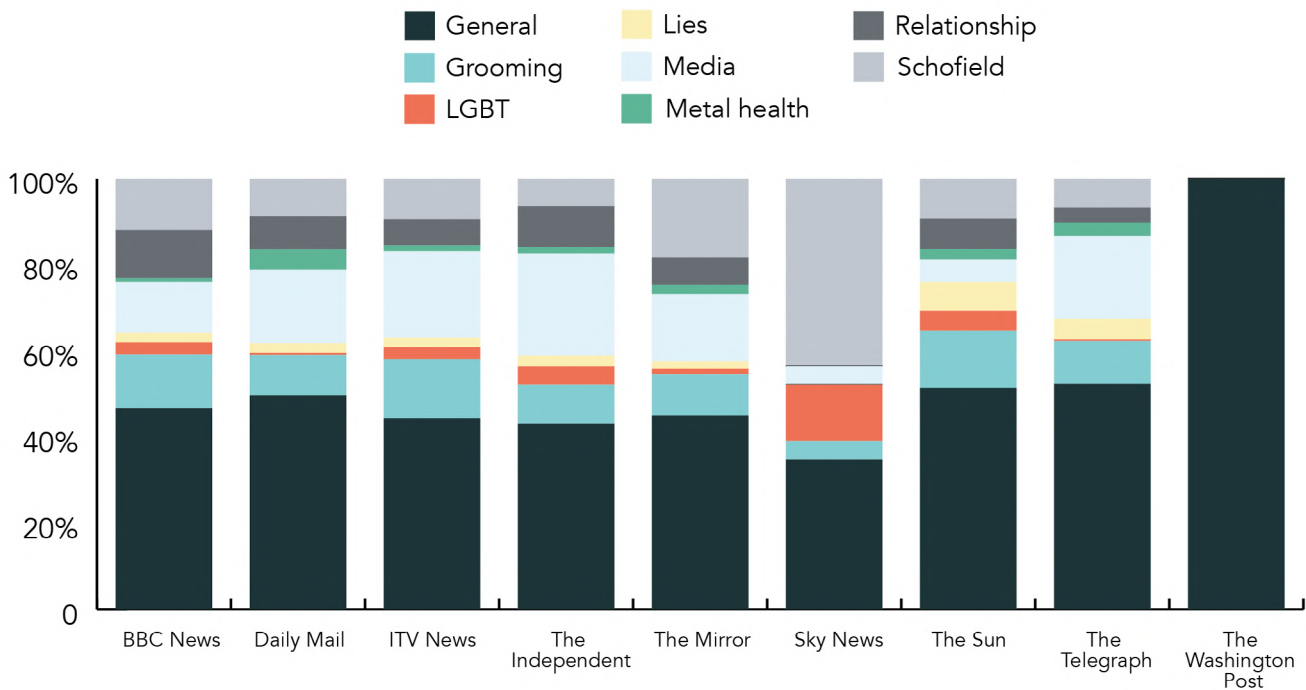


FIGURE 9

PERCENTAGE OF REPLIES TO POSTS RELATED TO PHILIP SCHOFIELD BY NEWS ORGANISATIONS WHICH CONTAINED EACH TOPIC



Again, the proportions of audience replies on each theme were fairly consistent across outlets and across different framings of the original news organisation post. (Those which are very different, Sky and the Washington Post, had few articles in this dataset).



KEY FINDINGS

Amidst a focus on controversy and scandal, news and social media become the subject of the story themselves

There are concerns amongst the media and the public that discourse is not healthy. Reporting and replies were self-reflective, discussing and critiquing the tone, focus or content of the discourse.

One of the key threads of the reporting was itself self-reflective, including commentary and analysis on the backlash itself, and who held what responsibilities in these situations.

Opinion pieces also put forward that the response to the situation should be more compassionate, and calling out the collective response to the situation as being driven by schadenfreude.

Headline: 'Phillip Schofield lied, cheated and failed... but enough is enough - show some compassion'

Headline: 'We loved the Phillip Schofield drama because we enjoy watching people suffer'

There was also coverage of both public and high profile individuals criticising the nature of the backlash. This included both the backlash from the public, the media and the actions of ITV itself, such as after an interview in which Schofield made a comparison to Caroline Flack, who died by suicide in 2020.³³

Headline: 'Fans call on Eamonn Holmes to stop 'bullying' Phillip Schofield'

Headline: 'Piers Morgan says Phillip Schofield is 'broken' and urges 'baying mob' to lay off'

Headline: 'Elton John defends Phillip Schofield and says affair reaction is homophobic'

There were also users engaging in counterspeech by calling out the hate against him, particularly relating to the impact on his mental health. This was not only commenting on the situation or offering support in general, but actively addressing the online commenting community as a whole to change their behaviour.

Comment: 'This is homophobia. And at it's worst'

Comment: 'What exactly did he do wrong? Apart from making a mistake that wasn't wise, having an affair with a 20-yr old younger man doesn't mean there should be all this hate against him'

Comment: 'Back off - let him find some equilibrium...I do feel sympathy - please, let us leave the man alone'

Comment: 'It is time he be left alone now. Caroline Flack - Remember, be kind!'

There was also a great deal of discussion about the media and the role of the media in reporting on the story, frequently negative. This included frequent criticism of the fact of the reporting - such as on the grounds of intrusion, or irrelevance:

Comment: 'Going to somebody's home - it's completely disgusting. How dare they be hounding a person on their doorstep. The media is disgusting'

Comment: 'Why is it that mainstream media thinks this Scofield story is this important? They seem like they think its > important than the migration crisis, the NHS is floundering, the Ukraine war, the economy...why??'

Comment: 'People are murdered, everyday, & it doesn't ever get mentioned bcos they aren't famous. Whilst this crap's been going on for several weeks'

33 <https://www.bbc.co.uk/news/uk-england-london-53676793>



Audiences are not bound by news media reporting standards and so news fuels speculation

Headline: *'I'm broken and ashamed... but I'm NOT a groomer says Phillip Schofield in bombshell first interview'*

A widely discussed feature of the relationship was the age gap between Schofield and the colleague with whom he had the affair. Posts on this topic generally centred either facts about the relationship or Phillip Schofield denying any accusations of 'grooming'; the replies often did not. There was significant online discourse using pejorative terms attacking Schofield and making false allegations against him, particularly linking him to his brother.

This also demonstrates the challenge of a content-moderation approach to tackling online harms. For instance, we found relatively little explicitly intentional homophobic language, (our hypothesis being that much of this content would have been successfully moderated out by the platform before our data collection took place.) However, the narratives surrounding the story clearly echoed homophobic narratives, and trends seen in gendered disinformation against LGBT people - but often disavowing any explicit connection and meaning that it would be much more difficult to say of a *specific* piece of content that it was homophobic, even though the overall narrative being amplified is one with homophobic undertones. There were many one-word replies (pejorative terms), with the short and rapid response facilitated by social media lending itself perfectly to amplifying a torrent of abusive comments.

However, there was also a community of users who were speaking out against the pile-on that was occurring by trying to persuade their fellow commenters to refrain from sharing hate or attacks, often out of concern for the target's wellbeing, as well as on general principles of justice.

Media organisations and personalities being at the heart of the story became a wedge for conspiratorial narratives to gain a toehold

There was deep distrust of media organisations, including allegations of collusion, corruption and criminal conspiracy. At the milder end of the spectrum, this discussion represented annoyance or legitimate critique, ranging up towards outlandish conspiracy theories, calling for media organisations to be shut down, and using mistrust of the media to promote disinformation.

Comment: *'ITV is a completely corrupt org, it plays the public as if they are fools'*

Comment: *'I do actually hope now that now we know how ITV is as well as the BBC, the whole darn lot will crash down: it's one big paedo ring'*

Comment: *'They've known all the time what he was doing, and encouraged this awful behaviour. We should be protesting, and get them shut down!'*

There were also more extreme conspiratorial narratives, particularly around support for Andrew Tate, and notions of a mainstream media 'conspiracy' against Tate. Andrew Tate is a widely popular misogynistic online influencer³⁴ (who has also been charged with rape and human trafficking).³⁵ He has popular online³⁶ with people alleging a conspiracy to silence Tate. In the Schofield discussion, we saw criticism of the BBC for alleged double standards, of not treating Schofield as negatively as Andrew Tate had been treated. The narrative being promoted in these posts was that Schofield had committed a wrongdoing whereas Tate was being falsely accused - amplifying the idea of a media conspiracy and using the story being discussed (which was unrelated to Andrew Tate) to defend Tate, a renowned misogynist.

However, there were only 36 replies which mentioned Andrew Tate, which is a small proportion, but nevertheless a concerning narrative.

Comment: *'Hold on, he's free, he came and said that he had done it, and nothing has happened to him? For seven months Andrew Tate's been fighting, when they've got zero proof, no evidence, and more than two hundred females said that he hadn't done anything to them - and as well as that, they want that we should feel sorry for him?'*

34 <https://www.theguardian.com/technology/2022/aug/06/andrew-tate-violent-misogynistic-world-of-tiktok-new-star>

35 <https://www.bbc.co.uk/news/world-europe-65959097>

36 <https://www.mediamatters.org/rumble/right-wing-influencers-and-media-figures-defend-andrew-tate-after-he-charged-human>



Comment: *'And people want Andrew Tate to be silenced - the news, it's as fake as it can get'*

There were also more general conspiracy theories mentioned, such as that surrounding the 'Great Reset', a theory which attributes the Covid-19 pandemic to the actions of an elite conspiracy:³⁷

Comment: *There's all of this garbage, old conspiracy theories which are really true, distracting you. At the same time the scum government are signing a pandemic treaty to push the Great Reset. More garbage out of the scum news corporations'*

Particularly notable is the presence of Andrew Tate defenders in these replies. Even when there is little connection between their stories, his supporters draw a parallel in order to try to discredit the mainstream media. A link is made between unhappiness at the BBC or ITV's actions relating to this particular event to a much wider conspiratorial viewpoint about the power of the media, championed by Tate's followers. Arguably, this clouds valid criticism of the media and makes the democratic discussion more difficult to have, for fear of playing into existing tropes about the 'MSM'.

These reflections indicate a level of disconnection between media and audiences, with some audiences participating in harassment of and conspiracising about media outlets, as well as parts of the media calling out audiences and other parts of the media.

However not all of the commentary about the media was negative. Some defended those working in media, and some commenters had ideas for improving the media ecosystem:

Comment: *'NRK's well known for its high-quality journalism - I do think TV's done well over there. If we had a national tax in the UK it would mean paying just a few £ a year, as opposed to the licence fee at the moment'*

These reflections indicate a level of disconnection between media and audiences, with some audiences participating in harassment of and conspiracising about media outlets, as well as parts of the media calling out audiences and other parts of the media.

³⁷ <https://www.bbc.co.uk/news/blogs-trending-57532368>

DYLAN MULVANEY

A CASE STUDY ON CORPORATIONS, TRANSPHOBIA AND ANTI-LGBT DISCOURSE

KEY FINDINGS:

1. Political debate and democratic expression ends up intertwined with divisive narratives in audience discussions of news.
2. Influencers, celebrities and politicians prolong controversies, taking their steer from online discourse and feeding further news coverage.
3. True, misleading and weaponised information are all used to promote harmful narratives.
4. Media is playing an important role in amplifying counterspeech, but this can lead to a backlash of further harmful discourse from audiences.

THE STORY

Dylan Mulvaney is a trans influencer who became famous on TikTok with her series 'Days of Girlhood' documenting her transition.³⁸ At time of writing, Mulvaney has 10.5 million followers on TikTok. In April 2023, the American beer brand Bud Light partnered with Mulvaney, sending her beer cans including cans with her picture on them, which she showed in a promotional video.³⁹ This sparked a backlash and a boycott of Bud Light by those who objected to the company partnering with a trans woman, as well as a wave of transphobia and harassment against Mulvaney, who experienced death threats.⁴⁰

This is within a wider context of ongoing debate in the US about corporate partnerships being criticised for being too 'woke' or aligned with social justice movements; or alternately, being criticised for not demonstrating genuine allyship. There has been a backlash against the retail store Target for stocking Pride items,⁴¹ as well as against other companies such as Nike, also for partnering with Dylan Mulvaney.⁴²

38 <https://365daysofgirlhood.com/>

39 <https://www.vox.com/money/2023/4/12/23680135/bud-light-boycott-dylan-mulvaney-travis-tritt-trans>

40 <https://www.nytimes.com/2023/06/29/business/bud-light-dylan-mulvaney.html>

41 <https://www.reuters.com/business/retail-consumer/target-pride-backlash-exposes-rainbow-capitalism-problem-designer-says-2023-05-31/>

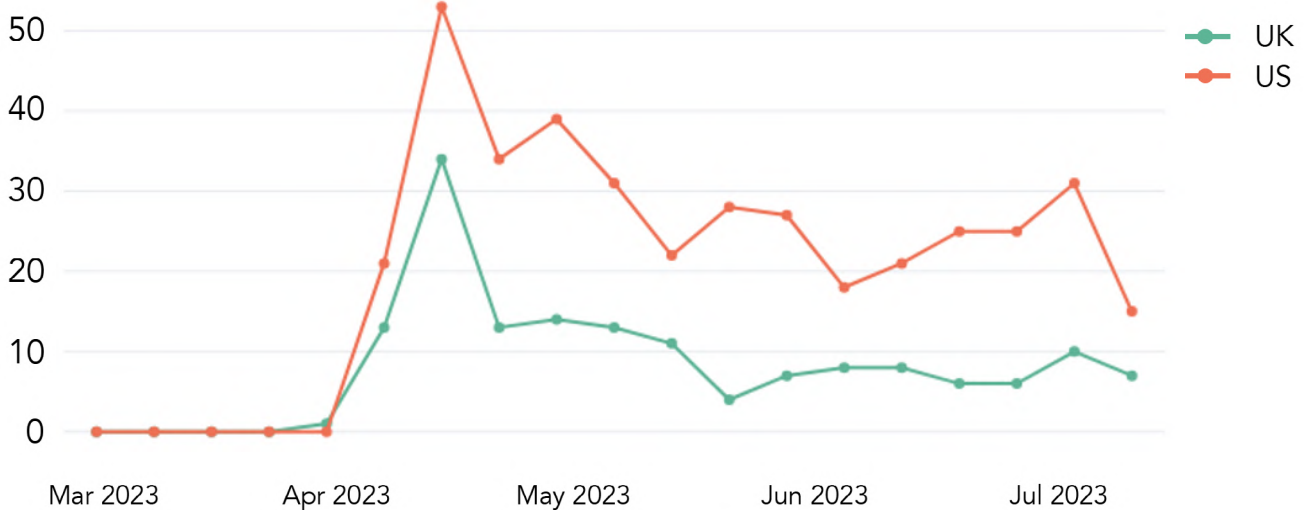
42 <https://www.dailymail.co.uk/news/article-11943777/Nike-angers-feminists-making-trans-influencer-Dylan-Mulvaney-new-face-womens-workout-gear.html>

HOW IT SPREAD ONLINE: ENGAGEMENT AND KEY THEMES OF DISCUSSION

Given that the influencer and corporations involved are US-based, this was primarily a US story, with more posts by right-leaning outlets overall. Most posts were posted by Fox News (262 posts in the timeframe); however it also gained some traction in the UK, with the most posts by the Daily Mail (130). The reporting of this story was highly engaged-with across the platforms we investigated.

FIGURE 10

VOLUME OF ARTICLES RELATED TO DYLAN MULVANEY OVER TIME, FROM US- VS UK-BASED NEWS ORGANISATIONS



News organisation coverage focused on both social justice and on the role of corporations (see table below). As most coverage was from right-leaning outlets, their focus on the impact on Bud Light and the role that ESG should play in corporate decision-making was reflected in the overall dataset. Left leaning outlets' coverage was more commonly focused on social justice, while centrist outlets focused on both social justice and the impact on Bud Light.

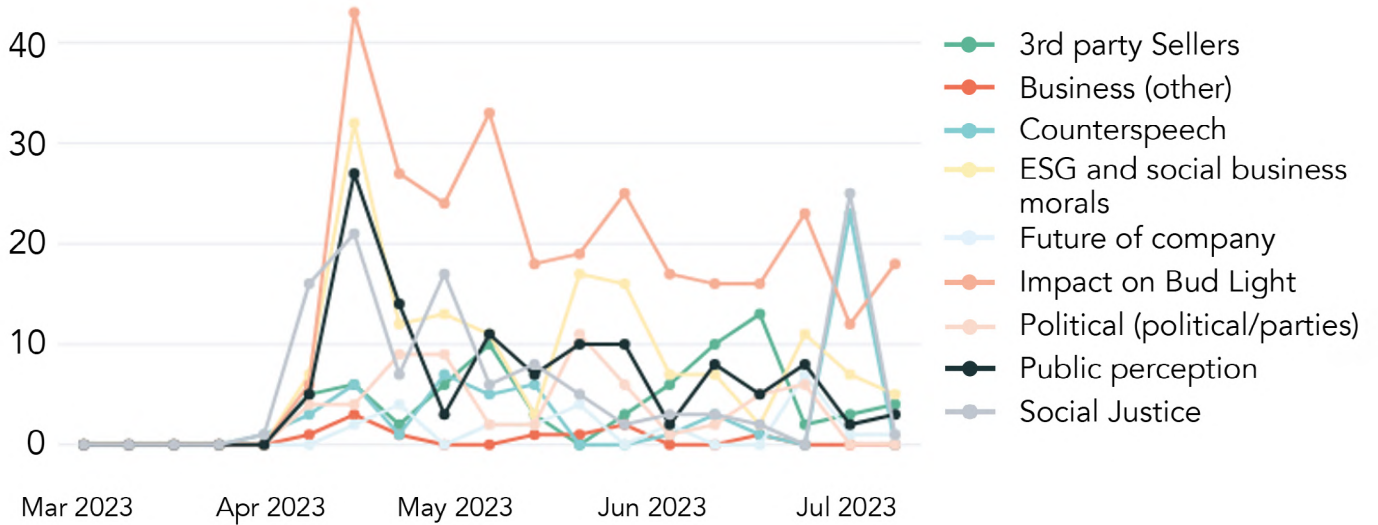
TABLE 6

MAIN NARRATIVES OF DISCUSSION RELATED TO DYLAN MULVANEY WITHIN OUR DATASET OF POSTS FROM NEWS ORGANISATIONS

NARRATIVE	DESCRIPTION	EXAMPLE HEADLINES	NUMBER OF POSTS
Relating to business: focusing specifically on the corporate element of the partnership, and what it means for what businesses should do	Impact of the controversy on Bud Light's business operations and finances, such as discussions and analysis of the impact on sales or on employees of the boycott of Bud Light	'KICKING THE CAN: New stats reveal just how badly recent backlash is hitting Bud Light's profits at the link in bio.'	303
	Future of Bud Light and Anheuser-Busch (Bud Light's parent company)	'Bud Light is 'coming back' but controversy is an 'important wake-up call,' Anheuser-Busch exec says'	25
	Actions taken by 3rd party sellers of Bud Light (e.g. participating in or avoiding the boycott)	'Boston Red Sox fans ditch Bud Light as Dylan Mulvaney backlash continues No one wanted to buy it 😞'	76
	Public perception of the controversy	'Bud Light salespeople receiving 'middle fingers' and 'car horns' as boycott passes seventh week: report'	117
	The role of corporations in promoting ESG principles and whether or not corporations should become involved in political issues	'Bud Light's inclusive ad campaigns are good for business'	152
	Involvement of politicians or political parties	'POUR ONE OUT: @[Senator username], GOP lawmakers sound off on Bud Light's plummeting sales.'	65
Relating to social justice: primary focus on discussion of transphobia or wider social justice issues	Social Justice subjects, both supportive and critical: e.g. LGBT rights, 'wokeism'	'Conservatives mocked for boycotting Budweiser for supporting trans rights'	124
	Counterspeech: speech highlighting criticism of harmful narratives, or amplifying the voices of those affected e.g. Mulvaney herself	'Zach Bryan Uses Travis Tritt Lyric to Condemn Transphobia Country Star Zach Bryan Quotes Travis Tritt Song as He Condemns Transphobia: 'Completely Wrong''	64

FIGURE 11

VOLUME OF ARTICLES OVER TIME WITHIN OUR DYLAN MULVANEY DATASET, SHOWING ALL MAIN NARRATIVES

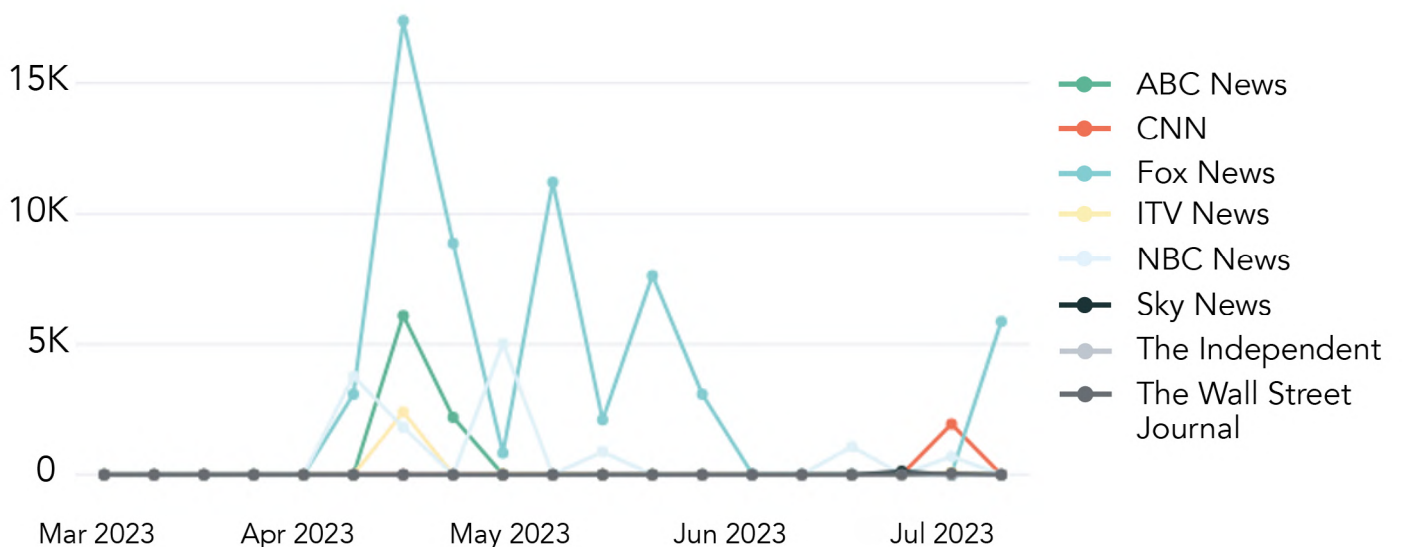


As with our other case studies, the vast majority of replies to these posts came from comments on YouTube videos, likely due to data access restrictions (see Limitations).

This also means that the replies we analysed in detail are skewed towards those outlets which tended to post more on YouTube - such as Fox News - although in this case, this was also the outlet which posted the most about the case study overall as well. Looking at the volume of replies, Fox News posts had the most replies overall - however they had the highest number of posts. If these figures (YouTube only) are normalised, meaning the disproportionate numbers of posts across different outlets are taken into account, Fox News and ABC News lead the replies.

FIGURE 12

VOLUME OF COMMENTS OVER TIME WITHIN OUR DYLAN MULVANEY DATASET, BY NEWS ORGANISATION



**TABLE 7**

MAIN NARRATIVES OF DISCUSSION WITHIN OUR DATASET OF USER REPLIES TO NEWS ORGANISATIONS' POSTS RELATED TO DYLAN MULVANEY

THEME	DEFINITION	EXAMPLE COMMENTS (BOWDLERISED)	NUMBER OF COMMENTS
Other	Replies where the theme was not strong enough to be identified by the topic model Includes various discussion of e.g. the role of the media, general political discussions, discussion of religion	'Yep, I knew it's a matter of time before the Right is crazy again. This is the exact response they wanted to see. All this hate speech started up, and now they can justify their censorship. As much as I cannot stand Democrats, this is the exact reason I'll never call myself a Republican.'	45053
Boycott	Discussing the boycott - often supporting, but also general discussion or critiquing	26 percent [reported drop in Bud Light sales] is not enough. We can do b3tter! This boycott isn't temporary, it is a full-on life time ban!	26620
Transphobia	Amplifying, participating in or calling out and critiquing transphobia	'There is a pint of transphobia around the internet, due to a fifty-gallon drum of the corporate media sexually indoctrinating kids	10829
Corporations	Discussing the appropriate role of corporations in promoting social values	'You're a company, you should promote the products, not ur political values! '	4099
Wokeness	Discussing 'wokeness': often anti-woke, but also including those affirming the value of 'wokeness'	'Don't embrace woke culture! It's a mental illness out of Hell'	2481
LGBT	Included significant homophobia against LGBT communities as well as counterspeech	'[username] You do not "have" to react to LGBT people. They're no threat to you, or anyone else. Live your life and let them live their own lives.'	629

FIGURE 13

PERCENTAGE OF REPLIES TO POSTS RELATED TO DYLAN MULVANEY BY NEWS ORGANISATIONS WHICH CONTAINED EACH TOPIC

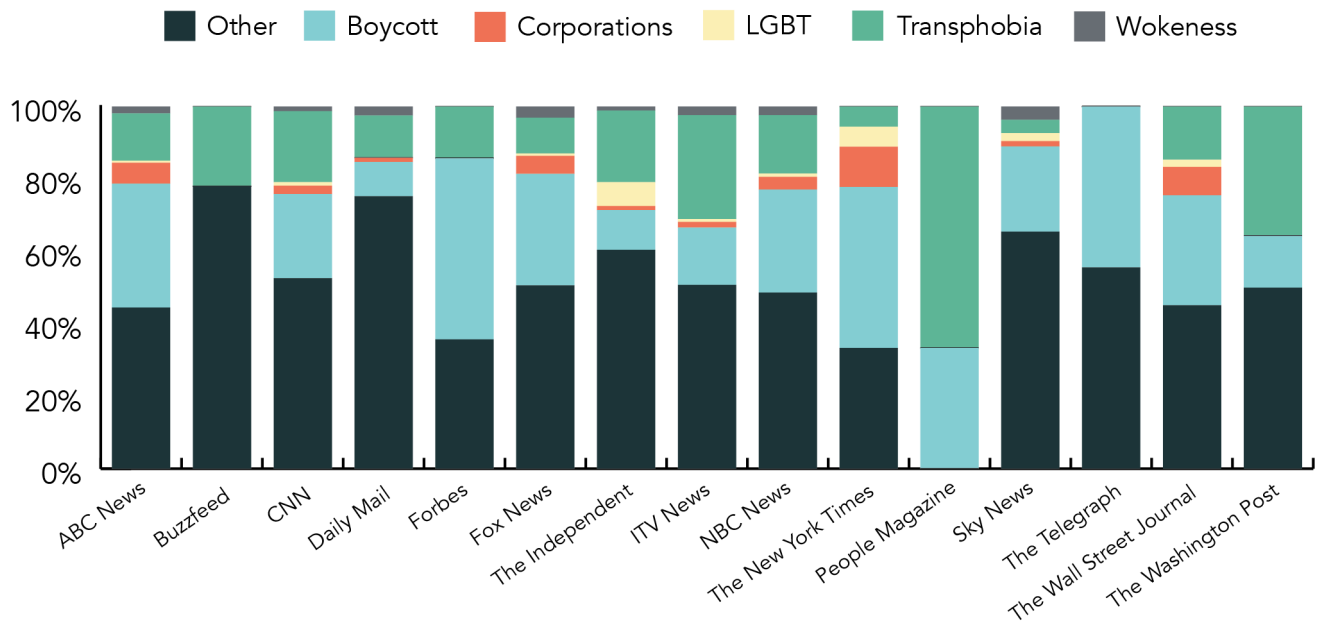
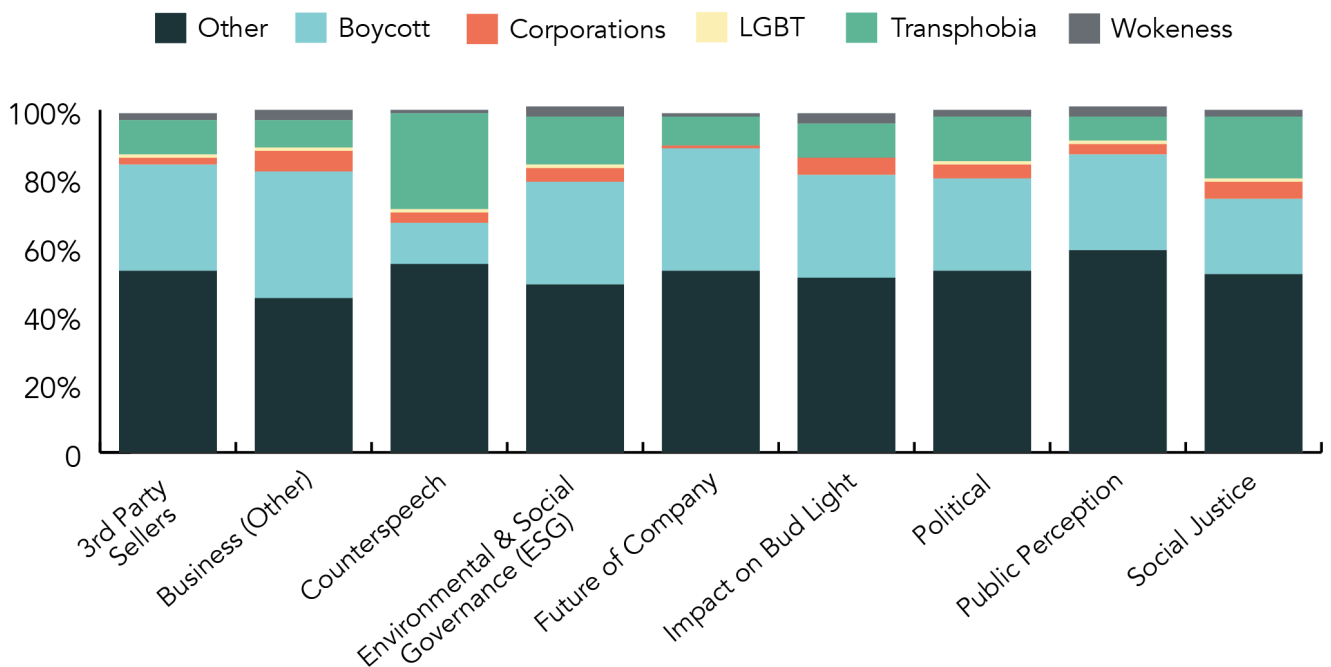


FIGURE 14

PERCENTAGE OF REPLIES TO POSTS RELATED TO DYLAN MULVANEY WHICH CONTAINED EACH TOPIC, BY NARRATIVE OF ORIGINAL POST





These graphs show the proportions of replies falling into each theme that was generated from original posts, broken down by the theme of the post, and broken down by the outlet which shared the original post. The proportions are fairly similar across the different outlets, and also across posts with different original themes, indicating that audience replies are relatively homogenous - extremely diverse in subject and tenor, but not radically shifting in response to who is posting about the issue or what is being posted. This suggests that rather than highly receptive or responsive audiences to differences in coverage, audiences are bringing their own narratives into the conversation: potentially, in the case of the more extreme or conspiratorial narratives, ceded in more alternative or radical spaces which then enter mainstream discourse.⁴³

KEY FINDINGS

Political debate and democratic expression ends up intertwined with divisive narratives in audience discussions of news

There are elements of free and pluralistic democratic discussion within this discourse - in particular, for instance, what approach corporations should take to ESG and on what basis they should make decisions about ESG, about which there were many different views expressed from a variety of standpoints.

The boycott theme was the most commonly expressed in replies to posts. There were many different views expressed from a variety of standpoints. Many commenters felt that corporations either should not be 'pandering', and should stay out of (seeming to be) taking a position on any political or identity-related issues; or that it was a strategically bad move by Bud Light; or not true or effective allyship to the LGBT community but merely a commercial move.

Comment: 'Companies can't ever be real allies. Their just in it for the \$ regardless of what 'side' they're on.'

Comment: 'I reckon by next year all marketing textbooks in college all over the USA will include a section on a company not understanding their market base'

These stories were often taken as indicative of companies succumbing to pressure from leftist or social justice movements, with the popular slogan⁴⁴ 'go woke, go broke' featuring frequently (1,081 times in the replies).

Comment: 'Woke's broke. When are they going to learn? How many folks are going to lose jobs since the company lost lots of money? People without any say in the marketing of products which they produce and package.'

Comment: 'Wokeness prevails bcos they've grabbed all the megaphones. Media, academics, entertainment - and more and more corporations...when a few make loads of noise with a platform from which to push an agenda or shame those who aren't complying. Corporations will take the easier path, giving in...'

Views against companies' support for the partnership included both criticism of the prioritisation of political statements over economic and social impacts, but also transphobia and conspiracism about the power or 'agenda' of the LGBT community and its allies.

Comment: 'The same people own all the corps: the ones who want 2 shape the world, in the deeds of the devil! That is why it is always being forced on everybody'

Comment: 'Yes, we have to do it more strongly, and boycott all the LGBT communities until they are bankrupt! It is only one percent of sick LGBT people and 99 percent of us who have healthy brains, we cannot risk it because of one percent'

Comment: 'LGBTQIA was no big deal, until they started coming after our kids, openly'

These two narratives - LGBT and Boycott - frequently overlapped in the audience responses. Rather than a discrete category of anti-LGBT discourse, political dissatisfaction (feelings of alienation, or lack of representation in corporations and the media) becomes intertwined with ascription of malign intent to other communities.

⁴³ This phenomenon was raised in a couple of our expert interviews.

⁴⁴ https://en.wikipedia.org/wiki/Go_woke,_go_broke



Influencers, celebrities and politicians prolong controversies, taking their steer from online discourse and feeding further news coverage

Rather than focusing on the partnership itself, news coverage tended to focus on the wide controversy around the partnership - covering the backlash, and then further responses to the backlash.

Influencer and celebrity responses were reported by news organisations: for instance, Kid Rock's response, who shot cases of Bud Light in protest of the partnership, was mentioned in 13 posts, including other celebrities' responses to Kid Rock.⁴⁵ Singer-songwriter Garth Brooks was mentioned 22 times for saying he would continue to stock Bud Light in his bar, which then provoked further backlash against that decision from other commentators.⁴⁶

Politicians also featured in news coverage, including Sarah Huckabee Sanders, Ralph Norman, JD Vance, criticising the partnership. News coverage not only focused on their critiques or commentary, but in some cases engaging in provocative activities which would appeal to a social media audience.

Headline: 'Sarah Huckabee Sanders trolls Bud Light with koozie for 'real women''

The controversy itself became the controversy. Online backlash to news reports then fed further news reporting, which itself led to further online backlash. This is a symbiotic relationship between news and consumer of sorts, but one which, in an engagement-driven environment, focuses on controversy and outrage.

True, misleading and weaponised information are all used to promote harmful narratives

There was also some evidence of gendered forms of disinformation. Gendered disinformation builds on tropes about gender to spread a particular narrative.⁴⁷ Less directly connected to the news story but in this category were replies focusing on studies which had come out about drops in male sperm count across America, warnings around the effects of 'gender-bending' chemicals in plastics,⁴⁸ and talking about birth control pills in the water supply, combined with transphobic attacks.

Comment: 'Its own transsexuality, mangina/eunuch/chemical-feminisation'

These narratives are an example of the different intersections between genuine news being distorted in online spaces to fuel conspiracism and hate. There have been regulatory interventions over the presence of BPA, an endocrine disruptor, in plastics. This then being represented as 'gender-bending', including in some reporting, along with conspiracy theories spread by public figures that hormones in water could turn people gay,⁴⁹ is combined in digital environments with gendered tropes about trans people and trans healthcare.

This results in discourse which weaponises, intentionally or unknowingly, different sources of information to support or 'rationalise' identity-based attacks against a particular group. Many replies did not link specifically back to the original story about Mulvaney or to transphobia outright. But as replies to posts about a trans woman's business partnerships, unrelated to discussion of endocrine disruptors, sperm counts or birth control pills, there is a clear throughline of the weaponisation of people's views around trans identities into a more extreme and conspiratorial form and using it as a focal point to spread harmful narratives by participating in mainstream conversation about a trans person.

Comment: 'When those in power say to kids that they are allowed to pick their gender and men are able to have a baby, it will be the end for the human race...Nature is, much cleverer than a group of people who are mentally ill'

45 <https://www.independent.co.uk/arts-entertainment/tv/news/kid-rock-bud-light-transgender-advert-b2318356.html>

46 <https://www.billboard.com/culture/pride/garth-brooks-bud-light-backlash-inclusivity-1235356774/>

47 <https://demos.co.uk/research/engendering-hate-the-contours-of-state-aligned-gendered-disinformation-online/>

48 <https://www.dailymail.co.uk/health/article-11992321/The-six-toxic-gender-bending-chemical-lurking-common-household-products.html>

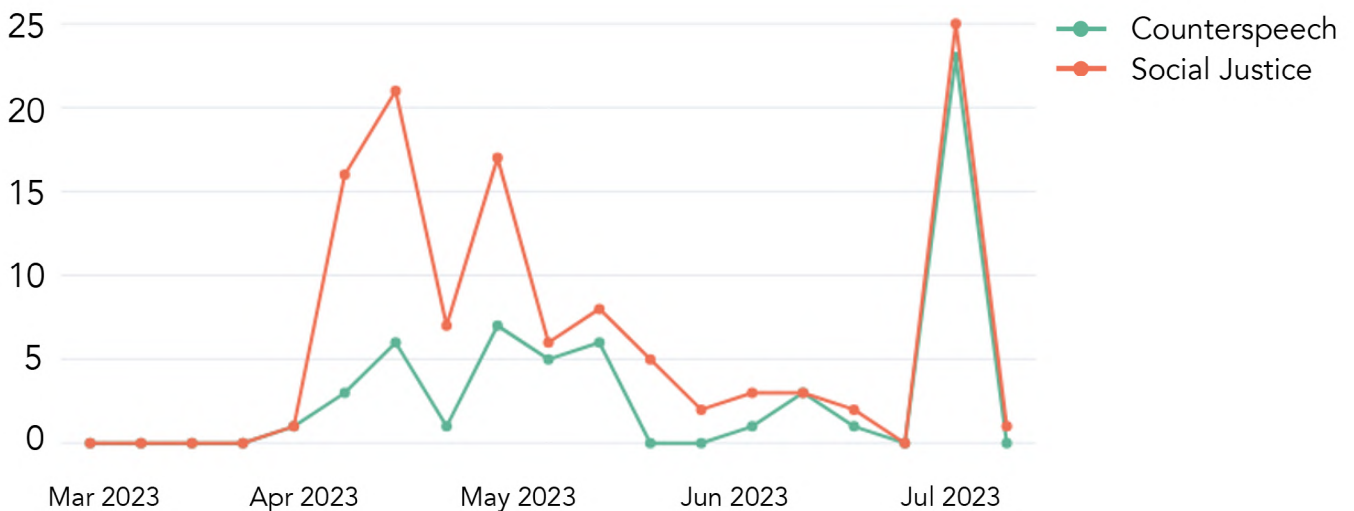
49 <https://www.rollingstone.com/politics/politics-news/desantis-disney-appointee-tap-water-turning-people-gay-1234690629/>



Media is playing an important role in amplifying counterspeech: but this can lead to a backlash of further harmful discourse from audiences

FIGURE 15

VOLUME OF ARTICLES OVER TIME RELATED TO EITHER COUNTERSPEECH OR SOCIAL JUSTICE NARRATIVES, WITHIN OUR DYLAN MULVANEY DATASET



The spike around July 2nd of posts which focus on counterspeech and social justice represents the news after Dylan Mulvaney spoke out about the lack of support she had received from Bud Light, the transphobia she had faced, the negative impact on her. This was a crucial way in which, in this discourse, the media played a role in amplifying counterspeech. Some outlets explicitly named transphobia in their headlines, and others highlighted quotes from her about her experience, including her statement that it was ‘worse than not hiring a trans person at all’ not to support a trans person in this situation.

Headline: ‘Dylan Mulvaney says Bud Light hasn’t supported her amid transphobic backlash’

However, there was significant transphobia in the resulting discourse, including discourse directed at Mulvaney herself and invalidating her identity or deliberately misgendering her, as well as personal attacks, and promoting transphobic narratives that being trans is a mental illness or is dangerous to society. It is worth noting that the media posts we reviewed gendered Mulvaney correctly, and indeed a failure to do so by CNN this year was widely called out.⁵⁰ Nevertheless, the replies often rejected this and often pointedly misgendered her.

These attacks included common narratives used to dehumanise and undermine LGBT rights - that they are not ‘normal’, or that LGBT rights are ‘dangerous’ and that measures should be taken.

Comment: ‘Freaks from the circus and people who are mentally ill are entertaining, but to treat them like we treat normal people is a dangerous thing’

Some replies were supportive or included counterspeech: people speaking up for or defending Dylan directly, or calling out transphobia by other commenters.

Comment: ‘Nope, this is about hate. If it really was about child protection, you would not be taking away the ability to get gender-affirming healthcare: loss of access to care will increase the rate of self-harm among young trans people’

50 <https://www.independent.co.uk/news/world/americas/dylan-mulvaney-cnn-apology-misgender-b2374527.html>



Others, however, defended transphobia itself:

Comment: *'It's not a problem for me to get called homophobic or transphobic. Yes, I'll wear it like a badge of honour! Period!'*

Comment: *'Not transphobic in any way, it's not normal. This is the usual news attempting to help push an agenda and make backlash against anybody who finds it not normal'*

Even though news coverage in this case platformed counterspeech, audiences did not always respond in kind, and in some cases doubled down on the harmful narratives that were being called out.

This case study demonstrates how even when 'good actors' online - such as news outlets - promote counterspeech, this is insufficient to address the problems of harmful speech and identity-based violence⁵¹ in response to controversies, in particular, transphobia and homophobia. Although not the entire discussion, the pervasive nature of false, divisive and harmful rhetoric means that these online spaces are far from the positive epistemic environment in which political discussions can safely be had.

51 <https://protectionapproaches.org/identity-based-violence>

ANALYSIS

LESSONS FROM OUR THREE CASE STUDIES

By examining these three cases in this way, we add detail and a new layer of understanding of what's going wrong in the information environment.

Audiences are key drivers of digital discourse, but this operates within a social-media engagement paradigm which does not lend itself to meaningful participation

- *Media is playing an important role in amplifying counterspeech: but this can lead to a backlash of further harmful discourse from audiences.*
- *Amidst a focus on controversy and scandal, news and social media become the subject of the story themselves.*
- *Influencers, celebrities and politicians prolong controversies, taking their steer from online discourse and feeding further news coverage.*

Although news organisations are frequently setting the subject of discussion on social media, in some cases they have less control over the tenor of the debate. Audiences are not passive recipients of the information they share, but engage actively in supporting or rejecting the framings they are offered. In our case studies, we saw that commenters frequently brought other topics into the discussion or rejected the framing offered in the original coverage (such as doubling down on transphobia in response to news posts amplifying its occurrence).

As in the discussions about Hunter Biden's various controversies, and the introduction of characters such as Andrew Tate into the discussion around Philip Schofield, we can see that audiences are bringing in their own scepticism about what the media are covering and why: what agenda it is part of, and linking into these discussions their own online communities and subcultures.

We also see how, across our case studies, news organisations and others who are influential in the public sphere respond to online discourse. Often, they play into its habits - we saw celebrities engaging in social-media friendly stunts which spark further reporting and discussion while politicians feed news coverage with soundbites; both often attack the reporting, adding a new layer to the dynamic. There is a cycle of controversy, backlash and outrage-provoking, which commonly drowns out constructive dialogue.

Democratic discussion is intertwined with division, hate, and false or misleading narratives, allowing these narratives to be amplified in online discussions

- *Political debate and democratic expression ends up intertwined with divisive narratives in audience discussions of news.*
- *True, misleading and weaponised information are all used to promote harmful narratives.*
- *Audiences are not bound by news media reporting standards and so news fuels speculation.*

Throughout our case studies, we saw engagement in free and pluralistic democratic debate and discussion. Audiences debated the appropriate role of corporations in promoting social values, called out wrongdoing by powerful figures, and exchanged views on how workplace relationships should be navigated to ensure safeguarding. However, these frequently overlapped with socially divisive and harmful speech and narratives that dehumanises, attacks or excludes others. We saw pejorative and cruel attacks on individuals, transphobia and homophobia, and the demonisation of marginalised communities, as well as the weaponisation of information to support divisive and false claims. This is far from the positive epistemic environment in which political discussions can usefully or safely be had.

Information chaos persists, and people are not satisfied with the quality of democratic discourse or information production

- Distrust in information and information institutions is endemic, and spans from legitimate democratic critique to outright conspiracism.
- Feeling unrepresented in mainstream media lends itself to conspiracism amongst audiences.
- Media organisations and personalities being at the heart of the story can become a wedge for conspiratorial narratives to gain a toehold.
- The distinction between news and opinion is blurry, as public figures' commentary frames factual reporting.

Political divisions over truth and disinformation, and the associated accusations of conspiracy, become the subject of news reporting and audience discussion, leading to further information chaos. The intertwined role of politicians and commentators in making headlines are also blurring the lines between what is commentary, opinion, and news reporting.

Amongst this chaotic environment, a recurring theme across our case studies, in particular that of Hunter Biden and Philip Schofield, was distrust, critique, conspiracising and disinformation about the media itself. There was widespread antagonism and hostility towards information producers and information institutions. We saw criticism of how media organisations reported on or handled controversies, up to accusations of corruption and grand conspiracy to hide the truth about politicians, pandemics or criminal enterprises from the public.

We also saw in the Schofield case study that reporting and replies were self-reflective: with media posts and public commenters both discussing and critiquing the tone, focus or content of the discourse, and expressing concerns that the discourse is not healthy.

In many ways, the findings of this research are further confirmation that the relationship between information producers and information consumers is broken. This is not simply a 'trust' crisis driven by any one player. It is a 'systematic trust crisis' in which mistrust, information chaos, information control, political alienation and commercial dominance drive each in turn.

If this is the problem, what does the positive vision look like? Next, we turn to experts from across the journalism industry, academia and civil society to understand what a healthy information ecosystem would look like - and what's preventing it from coming to fruition.

PART TWO

WHAT WOULD A HEALTHY INFORMATION ENVIRONMENT LOOK LIKE - AND WHAT'S PREVENTING IT?

Over the past two decades our information environments have changed beyond recognition from analogue to digital, from models in which people tended to buy into a narrow range of news brands, to consuming multiple distributed sources via social media. The world we set out above via our case studies would have been unthinkable in the analogue era.

In this section we examine what an optimal relationship between news, social media and citizens would look like. Based on interviews with a series of experts, we explore where there are tensions within this ideal, and how far away from it we are in reality.

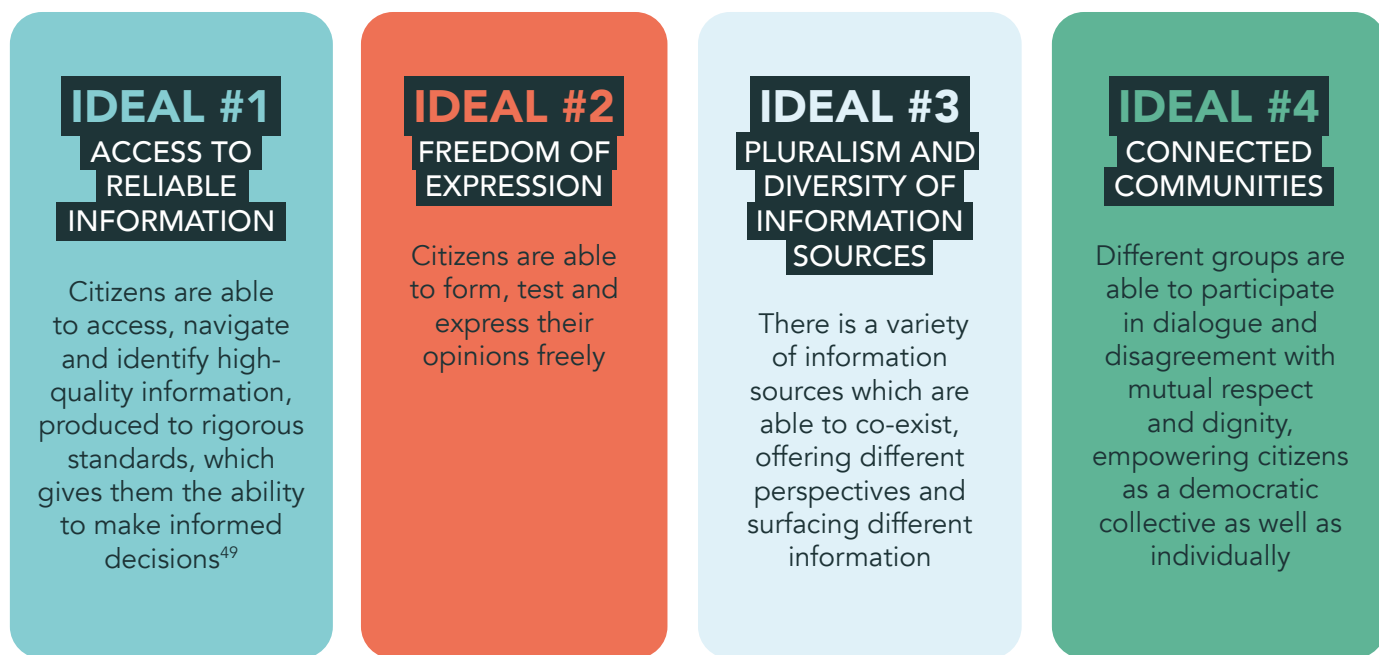
Among the experts we interviewed for this project, there was strong consensus that a healthy information ecosystem is the bedrock of a strong democracy, providing a base for informed, engaged citizens, who are able to hold elected officials to their promises and drive positive change in their own communities. Yet at the same time, many of our interviewees raised existential questions about the current and future role of journalists and traditional media within this broader ecosystem.

Building on existing evidence⁵², expert interviews and our wider programme of work into information environments, we set out four key aspects of a healthy information environment. When combined, these ideals mean that an information environment can function to fulfil core democratic purposes. These include enabling citizens to exercise their digital and human rights, and promoting healthy democratic discourse. In their absence, however, which is all too common, democratic discourse cannot be effectively sustained - as we have seen throughout crises such as conflicts, pandemics and political upheaval.

A healthy information ecosystem is one which enables connection, social cohesion, and protects citizens against violence, in which people are safe from abuse, and communities can organise and communicate. It should enable free expression - including the right to opinion and to receive information - to facilitate creativity, debate and discourse. There should be spaces for diverse opinions, and plural sources of information. And citizens should be able to access reliable information on which they can make informed decisions, be they personal or political.

52 <https://cdt.org/insights/from-our-fellows-envisioning-a-healthy-information-ecosystem/#:~:text=Healthy%20information%20ecosystems%20support%20diverse,%2C%20management%2C%20and%20securitization%20practices;https://misinfocon.com/designing-our-way-to-a-health-information-ecosystem-1efc97fe6000;https://newpublicsphere.stir.ac.uk>

Below we set out these four ideals and how they are being both realised and challenged in digital environments.



There are core tensions in achieving all four aspects at once. Policy interventions which prioritise one may often end up compromising another in its pursuit. For instance: free expression and pluralism of information can serve access to reliable information. Free expression is crucial for ensuring that truths and reliable information in the public interest can be shared and accessed widely, without suppression to serve narrow interests. Greater diversity of information sources improves knowledge creation and challenges hegemony.

But they can also conflict. Not all sources are reliable or acting in good faith, and can put greater burdens on citizens to be discriminating in their information consumption. Similarly, free expression in a pluralistic environment benefits communities to be able to communicate, organise and amplify historically marginalised voices. But it also allows harmful and divisive narratives which threaten the rights of particular groups to participate and find new audiences.

We aim to set out a diagnosis of the state of each ideal, and how news media, social media and citizens are interacting to uphold or undermine it.

HOW HAS DIGITALISATION AFFECTED THE REALISATION OF THESE FOUR VALUES?

Digitalisation has changed how these ideals manifest in our information environments - vastly improving some aspects, while introducing new threats to others. Below, we set out the four key trends that emerged from our interviews and evidence review that are inhibiting the realisation of these ideals. By doing so we set the scene for our subsequent policy recommendations in part three.

IDEAL ONE: ACCESS TO INFORMATION

Under this ideal, citizens are able to access, navigate and identify high-quality information, produced to rigorous standards, which gives them the ability to make informed decisions.⁵⁴ In our research and interviews we heard how the role of journalism has changed through the digital era in this respect. The quality and standards of journalistic content means that news organisations are in a strong position to act as information 'gatekeepers' to ensure citizens are accessing high quality information. Under this older paradigm,

53 <https://www.unesco.org/en/right-information>

54 <https://newpublicsphere.stir.ac.uk/final-report>

journalism's adherence to rigorous standards grants it the authority to determine in the public sphere what qualifies as reliable information. In the digital age and under the engagement-based paradigm of social media, this role is challenged by new actors and new forms of information control, while citizens are also widely questioning the authority of 'gatekeepers'.

We heard that acknowledging the subjectivity involved in news reporting and engaging citizens to help guide them in navigating information is a way to reconcile this challenge. The news media is struggling to retain its analogue role as 'gatekeeper', but it could shift its role to 'guiding' and helping citizens access and assess reliable information.

The news media's role has evolved with each technological revolution. With the industrialisation of print in the 19th Century and the creation of mass media, the media landscape shifted from peer-to-peer and 'conversational', to 'top-down and centralised' (Professor Jeff Jarvis, Director of the Tow-Knight Center for Entrepreneurial Journalism at the City University of New York). The advent of mass media instituted journalism as the primary route for citizens to access reliable information; however, with the rise of the internet, media has once again become more conversational, introducing new pathways to knowledge.

Nic Newman, Senior Research Associate at the Reuters Institute for the Study of Journalism, states that 'the story of the last 15 years' in journalism has been one of the disruption caused by digital media and social networks, which have diluted the 'core gatekeeper role' of the press. Now, anyone is able to share information with a very wide audience, expanding the available sources and types of news. Charlotte Dewar, Chief Executive of the Independent Press Standards Organisation, argued that there should be 'no apology for social media being an incredible source of information'.

However, with increased access to information, comes increased access to low-quality information. Dewar warned that in parts of our current information ecosystem, conspiracies and misinformation proliferate,⁵⁵ with potential harms to important areas such as public health.⁵⁶ Eliot Higgins, founder of Bellingcat, was concerned that as the boundaries between sources of information break down, reliable information may be more difficult to find, leaving an opening for malicious actors and foreign influence campaigns to take advantage.⁵⁷

Some of our interviewees argued that the risk of these harms demonstrated the essential function of journalism, which should be, as Lexie Kirkconnell-Kawana, CEO of the independent press regulator IMPRESS argued, to curate information and establish the veracity of information. In this view, a journalist has a public mandate to say 'I'm a person who can be trusted'.⁵⁸

Meera Selva, Chief Executive Officer of Internews Europe, believes that by maintaining rigorous editorial standards, journalism plays a vital role in the information ecosystem by ensuring that high-quality, factual information is presented to the public in an accessible way. From this vantage point, maintaining and rebuilding trust in news and journalism is a crucial path to a better information ecosystem where citizens are more easily able to identify and access reliable information.

There have also been calls not just for journalists to reclaim their role as gate-keepers, but also for social media platforms to adopt it. Such calls have become particularly acute in the case of information crises, to support access to reliable information through greater moderation, as well as introducing 'information centres'⁵⁹ where authoritative information (such as that produced by health authorities about Covid-19) can be easily and clearly accessed.

Professor Jarvis takes the view that trying to 'play whack-a-mole' with disinformation has been fruitless. To tackle this problem, we need new 'mechanisms' and 'institutions' to focus on finding and elevating 'good information'. However, ultimately, he believes that individuals 'have to be responsible for their own judgement'. Governments, companies and news organisations cannot, and should not, 'control' public discourse.

And, as both Selva of Internews and Guilherme Canela of UNESCO noted, the effect of gatekeeping - an institution, organisation or individual determining for the rest of the ecosystem what information is reliable or authoritative - can be positive or negative, depending on the gatekeeper. In the context of a global

55 <https://www.npr.org/2023/03/31/1166649732/conspiracy-theory-eating-bugs-4chan>

56 <https://committees.parliament.uk/oralevidence/12788/pdf/>

57 <https://pen.org/report/hard-news-journalists-and-the-threat-of-disinformation/>

58 This view is particularly pertinent to the debate around the Online Safety Bill and Digital Services Act as to whether media and journalists ought to be exempt from the scope of platform content moderation duties.

59 https://www.facebook.com/coronavirus_info/

information ecosystem, selecting trusted regulators requires multilateral collaboration between both states and big tech companies.

James Ball, an independent journalist and author, describes the traditional mode of journalism in which journalists are presented as objective arbiters of truth, who have no personal investment or background informing their work, which this expert considers to be a notion which is out of date. In a world where we recognise that not everyone should indeed be trusted, Professor Charlie Beckett, Director of Polis and the Polis/LSE JournalismAI project, argued that polls showing declines in 'trust in news' are at best 'irrelevant', and at worst are stoking unnecessary fear. Instead, journalists should focus on the quality of their work and its relevance to the lives of the public. Indeed, there are indications that trust in news in the US and UK may be rebuilding itself, even though still at low levels.⁶⁰

And social media has introduced new gatekeepers. While our interviewees noted that the internet was in some ways a democratic revolution, in many respects the online world contains the same limitations as the offline world, from the risk of further concentration of power in the hands of states or corporations,⁶¹ to design choices with significant unintended consequences. For instance, social media sites may choose to 'shadowban', limit the reach of,⁶² or remove various kinds of content altogether.⁶³ This can impact both news organisations and ordinary citizens.

Moreover, gatekeeping information can be dangerous: for some audiences, top-down efforts from governments and regulators to verify or 'fact-check' information are received positively, but for Higgins, there are concerns over how such measures could be abused by authoritarian states,⁶⁴ or even democracies.^{65,66} One of the concerns about the UK's proposed media exemption and protections for journalistic content in the Online Safety Act, is that it would potentially mean less moderation of content from news publishers regardless of the quality of the content or even (given a broad definition) the veracity of the publisher, providing a way for bad actors, such as extremists, to claim to be producing 'journalism' in order to claim privileges for their content.⁶⁷ This year has also seen a prime example of when the 'gatekeeper' role is abused, with a major lawsuit against Fox News in the US for defamation on the grounds of Fox's coverage claiming that the 2020 election was 'stolen' and knowingly amplifying false information about the election and Dominion voting machines to their audiences.⁶⁸

If the gatekeeping role of the news media isn't viable in the digital age, what's an alternative approach?

Rather than doing away with the role of gatekeepers entirely, some of our interviewees recommended an evolution of the role that reflects more one of a guide that retains some core and important aspects while shedding others. This would retain journalism's key role in editing information and ensuring high-quality standards are met in producing and presenting information, but rather than this being taken on trust, also being transparent about the reasons behind and processes of coming to those decisions, and acknowledging the subjectivity therein. Rather than presenting information as objective fact, by being honest about their inevitable political perspectives, and presenting the evidence for their conclusions, trust can be built with audiences as they are allowed to make up their own minds. Despite the difficulties of a less hierarchical information ecosystem, both Selva and Higgins stated that the most effective way to tackle misinformation and disinformation is not through top-down (state/government) interventions, but rather from high-quality independent media, investigation and analysis.

Further, Higgins argues that when trying to prevent the spread of misinformation or conspiracy theories, we are often treating the symptoms rather than the cause. The root of mistrust in media, even to the point of conspiratorial thinking, is a scepticism of sources of authority, including the government, media and medical professionals, which is often a legitimate response to a previous traumatic moral injury.

Social media's positive role of greater access and representation needs tempering with journalism's positive role. The role of journalism in providing citizens with access to reliable information must be protected; but in an age of information disorder, this may mean moving away from an assumption of trust. By building

60 https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2023-06/Digital_News_Report_2023.pdf

61 <https://pen.org/report/splintered-speech-digital-sovereignty-and-the-future-of-the-internet/>

62 <https://www.poynter.org/commentary/2022/journalism-big-tech-relationship/>

63 <https://www.theatlantic.com/ideas/archive/2022/12/twitter-files-content-moderation-transparency/672468/>

64 <https://pen.org/report/splintered-speech-digital-sovereignty-and-the-future-of-the-internet/>

65 <https://www.cato.org/policy-analysis/jawboning-against-speech>

66 <https://rsf.org/en/uk-national-security-bill-poses-alarming-threats-journalism-and-press-freedom>

67 https://demos.co.uk/wp-content/uploads/2022/05/Exemptions_Exceptions_and_Exclusions__OSB__vF.pdf

68 <https://www.theguardian.com/us-news/2023/apr/18/fox-dominion-settle-us-defamation-lawsuit>

new investigative communities, giving citizens the tools to be able to investigate what is going on in their communities, people are able to verify information for themselves: these networks, connections, and access to information can empower the public. Restoring a sense of trust in established authorities may be less desirable than having a critical, well-informed public.

IDEAL TWO: FREEDOM OF EXPRESSION

Freedom of expression is a foundational pillar of a healthy information ecosystem, encompassing the ability to hold, share and distribute opinions and information, as well as the right to seek and receive a broad range of information.⁶⁹ A healthy information ecosystem also empowers information consumers, amplifying their voices into information producers.

The amount that digital environments have enabled freedom of expression should not be understated. New modes of communication means that citizens have a huge array of ways and spaces in which to form opinions, engage in debate and express themselves. This includes independent journalists and emerging media organisations who have access to new data and new audiences for their work. But at the same time, harmful content at scale can drive people away from engaging safely and freely in discourse - particularly members of marginalised groups who are more likely to suffer pile-on abuse and harassment, or be targeted in disinformation campaigns. Tackling the ways in which online spaces are built and managed is needed to reduce the risk of these harms without clamping down on widespread speech.

Across our interviews, our experts agreed on the necessity of a free press for a strong democratic process, generating both productive debate and reaching areas of consensus.⁷⁰ The shift from an analogue to an online information ecosystem has also had a democratising effect, enabling individuals to reach an audience in ways which were not possible through radio or broadcasting - substantially changing the power balance between public and press. Selva, Heawood and Professor Jarvis see the increased decentralisation of control over mass media as a positive, reducing the barriers to entry into the news market and enabling voices that were previously excluded in public discourse to be heard: from the formation of informal communities such as 'Black Twitter', to outlets built by-and-for minority communities, such as gal-dem.⁷¹ Overall, a far broader range of citizens are now able to drive public discourse and opinion, presenting traditional journalists with opportunities to listen and collaborate.

Canela described the expansion of the online information ecosystem as the 'biggest positive revolution for freedom of expression since Gutenberg', highlighting the positives of the internet and social media as a force for democratisation in the production and distribution of information. In this way, social media functions as a kind of universal platform, where anyone can express their own viewpoint, across sites and mediums (text, audio, visual, video).⁷² However, historically marginalised groups are still the most subject to online suppression in various forms.

But alongside the mass democratisation of the media, new threats have evolved: disinformation and radicalisation that disempower citizens, by contributing to information pollution, or by excluding them from being able to safely participate in these ecosystems at all.

Canela went on to state that while disinformation is 'as old as truth', new technologies have drastically increased the volume, velocity and potential virality of both disinformation and hate speech, with higher quality deepfakes a particularly worrying example, which even extensive media literacy efforts may fail to tackle.

There is real damage caused by radicalisation. There is evidence that extreme ideas can filter from fringe sites into the mainstream, with an expansion of toxic and violent rhetoric, with significantly negative impacts for the participation of marginalised groups in public discourse.⁷³ In the case of the Covid-19 vaccines, Higgins states that individuals who were worried or unsure about receiving the vaccine could easily encounter conspiratorial communities through search engines (such as when searching 'are vaccines safe'). While the vast majority of people are unlikely to take these conspiracies seriously, a small fraction will join these communities, and once

69 <https://www.ohchr.org/en/instruments-mechanisms/instruments/international-covenant-civil-and-political-rights#:~:text=Article%2019,-1.&text=Everyone%20shall%20have%20the%20right,3>.

70 <https://committees.parliament.uk/publications/6878/documents/72529/default/>

71 <https://gal-dem.com/>

72 https://knightfoundation.org/wp-content/uploads/2020/06/KnightFoundation_Panel6-Teclash2_rprt_061220-v2_es-1.pdf

73 <https://demos.co.uk/project/engendering-hate-the-contours-of-state-aligned-gendered-disinformation-online/>

surrounded by people who share a given belief, can become radicalised.

In the worst cases, the interaction online between different parts of the information ecosystem may further entrench existing power imbalances. Several of our interviewees highlighted the example of how women and ethnic minority journalists are targeted on social media platforms, to the detriment of their mental and physical health and safety, even to the extent of being harassed out of public conversation. In the midst of an online mob,⁷⁴ many justify their harassment as an expression of justice against the powerful, with little consideration of context or whether the individual truly is a powerful public figure. Ball highlighted how social media features such as Quote Tweets can be used to amplify harassment, abuse and disinformation,⁷⁵ with disproportionate impacts on women and ethnic minorities.⁷⁶ This can undermine the participation of marginalised groups in public life and democratic processes, as they are trolled out of the conversation.⁷⁷ At the same time, mandating the takedown of particular accounts can be easily abused to silence particular viewpoints. In online spaces, journalists simultaneously are overmoderated and underprotected: the same safety processes that mean their own content gets falsely labelled as state propaganda and blocked fail to protect them from harassment and threats.⁷⁸

The challenge in policy terms is how to balance avoiding interference in speech, while tackling the very real speech-suppressing harms that can result. The UK has decided to take a more liberal approach in the Online Safety Act, deferring the decisions of what to do about 'legal but harmful' speech to online platforms to set and enforce their own terms of service, with public pressure being the lever left to compel platforms to make changes which protect their users from resulting harms. The task ahead for the regulators in the UK will be how well they can address the platforms' failure to enforce these terms and conditions. Challenging the status quo will require considering how moderation decisions are made, enforced, or redressed. There are conceptual knots which have yet to be untangled about how our concepts of free speech apply in a digital era, where private companies provide the spaces and means of communication, and where 'freedom of reach' is something that can be suppressed or amplified by those companies. Regulation such as the Online Safety Act will need to successfully focus on how online spaces are built and managed and where these increase or reduce risks to users rather than taking a content-focused approach in order to combat these challenges.⁷⁹

IDEAL THREE: PLURALISM

Digitalisation offers a significant opportunity for amplifying a diverse range of voices and perspectives, by introducing new mediums and platforms. However, our interviewees raised how despite this, diversity and pluralism are not being achieved across the industry. From lack of diverse representation in newsrooms to the challenges of financial sustainability of emerging organisations and outlets, pluralism is difficult to scale and maintain. Audience engagement is dependent on the design choices of dominant technology companies about how news is treated on their services, and this poses a challenge to pluralism, as power remains concentrated rather than distributed across the information ecosystem. Competition regulation is an opportunity to work to rebalance this divide.

Canela described freedom of expression as allowing for the right to share information, but also the right to seek and receive ideas, perspectives, and entertainment. This generates and necessitates a pluralistic information ecosystem, containing a broad range of sources of information and varied perspectives.⁸⁰ In the view of Ball and Professor Beckett, the rise of online media has arguably created a far more pluralistic information ecosystem, allowing a far broader and more diverse range of perspectives and people to garner mass attention.⁸¹ Professor Beckett argued that encountering differences in opinion is a requirement for both freedom of expression and pluralism,⁸² even when it is unpleasant, and that the vast majority of online discourse is either 'sensible', 'boring', or 'pleasant'. On the other hand, some interviewees worried about a highly pluralistic information environment increasing the potential for extreme ideas to travel into mainstream discourse, which may itself then lead to a reduction in pluralism as some marginalised individuals cease to participate in public discourse.

74 <https://knightfoundation.org/reports/media-and-democracy/>

75 <https://edri.org/wp-content/uploads/2022/01/A-Policy-Guide-for-Protecting-Human-Rights.pdf>

76 <https://www.indexoncensorship.org/2022/11/where-are-all-the-voices/>

77 <https://pen.org/report/no-excuse-for-abuse/>

78 YouTube video

79 <https://demos.co.uk/blogs/system-change-for-system-changes-sake/>

80 <https://www.ofcom.org.uk/research-and-data/multi-sector-research/media-plurality>

81 <https://www.theguardian.com/media/2023/may/07/the-digital-media-bubble-has-burst-where-does-the-industry-go-from-here>

82 <https://knightfoundation.org/reports/media-and-democracy/>

Online media has the potential to cover stories which may not be surfaced by the mainstream press, whether by catering to specific niche audiences, developing new methods of in-depth reporting such as OSINT (open source intelligence investigations), or by using social media platforms to develop new modes of communication and reach previously underserved audiences.

However, the platformisation of journalism has also had negative effects on pluralism in other aspects of the information ecosystem.⁸³ Declining revenues have led to increasing competition for advertising and attention, with the gap between the resources for reporting at large, established outlets and smaller outlets widening substantially. In the United States, Dr Nikki Usher⁸⁴ argues that there has been a shift in the composition of legacy newsrooms, with a narrowing of the kinds of people producing news, and a subsequent shift in the audience - more educated and wealthy (rich), more racially homogenous (white) and overwhelmingly more likely to support the Democratic party (blue). This predominance of centre-left viewpoints among mainstream publications has alienated many more conservative US citizens, leading them to seek out more fringe, marginal and often extreme outlets. Social media is also frequently criticised for filtering people into 'echo chambers' through their algorithmic recommendation systems, where they are more likely to hear news that they like or agree with,⁸⁵ although the level of impact of this as a phenomenon is contested.⁸⁶

Even when many plural sources of information can be accessed and disseminated through digital platforms, if digital platforms are controlled by a few corporations which set the standards for information amplification and recommendation, that pluralism is seriously at risk. We have seen powerful examples in the disputes over news media bargaining code legislation in Australia and Canada. In these cases, the large platforms responded initially by blocking access to news content links on their sites: effectively cutting off one of citizens' most-used routes to access news and information, and for some smaller media outlets, cutting off their main route to access their consumers.⁸⁷

Similarly: when platforms control which kinds of content should be served in response to search queries or amplified in event of a crisis, it is likely that the established incumbents will be those to benefit most. Forthcoming competition legislation may be a particular opportunity to identify new interventions that could help rebalance power in the information ecosystem.

IDEAL FOUR: CONNECTED COMMUNITIES

The pressures on local journalism in a digital environment that thrives on scale and engagement has meant that digital spaces have become new sites of community news and information sharing. While these offer new opportunities for community engagement and connection, operating within the same dominant social media platform models means that similar amplification of harms and misinformation can occur, undermining community cohesion. There is a need for greater support to sustain local and public interest journalism to support communities at the local level in navigating information environments.

While often under-emphasised, one of the crucial democratic functions of a well-functioning information ecosystem is to enable the participation not just of individuals, but also of communities in driving discourse. When communities are able to communicate well and create social cohesion, they are able to make collective decisions, advocate together and be heard by their elected representatives, or make change directly, as opposed to isolated individuals all advocating for themselves.

Well-functioning local journalism is a particularly good example of this kind of positive force for collective action. When sufficiently funded and supported by community members, local outlets can focus on public interest stories and concerns close to home, which mainstream outlets are unlikely to be interested in or aware of. The cultivation of community in the information ecosystem does not have to be exclusively geographically-oriented, with social media communities and online news often created by and for marginalised groups, or other groups with shared interests, who can come together to create digital spaces and shared resources, from Freecycle⁸⁸ to Wikipedia.⁸⁹

83 <https://techpolicy.press/journalism-platforms-and-the-challenges-of-public-policy/>

84 <https://www.niemanlab.org/2021/07/white-audiences-who-will-pay-is-still-metro-newspapers-survival-strategy/>

85 https://assets.mofoprod.net/network/documents/Mozilla_YouTube_Regrets_Report.pdf

86 <https://techpolicy.press/from-filter-bubbles-echo-chambers-and-rabbit-holes-to-feedback-loops/>

87 <https://www.smh.com.au/national/a-pox-on-facebook-but-also-on-the-media-bargaining-code-20210218-p573qb.html>

88 <https://newpublic.org/directory>

89 <https://www.wikimedia.org/>

However, in our current information ecosystem, funding pressures have led to a reduction in higher-quality investigative, public interest or local journalism.^{90,91} Heawood highlighted increased incentives towards scale and engagement, either positive or negative, leading to lower-quality, high-volume clickbait,^{92,93} while Kirkconnell-Kawana drew attention to the increased prominence of local Facebook groups filling the gap left by local news. While these groups may serve some of the same functions of building cohesive communities, there is often little to no moderation, and when conflicts arise, can quite quickly devolve into a 'witch hunt' driven by 'mob mentality'. These communities are often private and difficult or impossible for researchers to ethically access, meaning that the role of these groups as micro-ecosystems of information sharing and connection are poorly understood.

Canela lamented that while the promise of the internet was many-to-many communication, enabling a plurality of senders and receivers interacting, we often have a few talking to too many, and too many talking to nothing or nobody. Too often we see the online information ecosystem reinforcing existing power imbalances,⁹⁴ with discourse driven by bad actors, towards divisive and even abusive ends.^{95,96} This process stokes division and hatred, undermining community wellbeing and social cohesion. To tackle these negative impacts of our online information ecosystem, there is a need for greater support to sustain local and public interest journalism to support communities in engaging with information.

90 <https://www.mediareform.org.uk/blog/mrc-submission-to-dcms-inquiry-on-sustainability-of-local-journalism>

91 <https://pen.org/wp-content/uploads/2019/11/Losing-the-News-The-Decimation-of-Local-Journalism-and-the-Search-for-Solutions-Report.pdf>

92 <https://www.gmfus.org/news/full-stack-approach-public-media-united-states>

93 https://www.freepress.net/sites/default/files/2022-03/to_protect_democracy_recreate_local_news_media_final.pdf

94 <https://www.isdglobal.org/explainers/the-groomer-slur/>

95 <https://demos.co.uk/project/silence-woman-an-investigation-into-gendered-attacks-online/>

96 <https://demos.co.uk/blog/gendered-disinformation/>

PART THREE

POLICY RECOMMENDATIONS AND A PATHWAY FOR CHANGE

Public interest news is having to compete for audiences in spaces not built for it. The production and consumption of information is shaped by a digital ecosystem which is based on advertising technologies and so prioritises engagement above all else. The result is that our public spaces for democratic discourse are not democratically healthy. We call this status quo the engagement-based paradigm.

The engagement-based paradigm might serve social media, it might even drive traffic to news media websites, but it is also polluting and corrupting our information environments. Through the detailed tracking of three online storms - the stories of Dylan Mulvaney, Hunter Biden and Philip Schofield - we have learnt more about how discourse evolves online. Our case studies reveal new details about how this paradigm is damaging discourse, amplifying potentially harmful speech and commentary, and poisoning our discourse - often to the detriment of minority groups. In the case of Dylan Mulvaney, we saw how the response to the story becomes the focus of debate in itself. In the case of Philip Schofield, we witnessed criticisms of the TV presenter's actions spiral into conspiracies about news organisations. Socially divisive and harmful speech and narratives that dehumanise, attack or exclude others cannot be conducive to a better and more democratic information ecosystem.

Our expert interviews give further insights. We heard that citizens need access to good information in order to engage in democracy, freedom of expression, pluralism and diversity of information sources and ways to connect as communities. But the algorithmic-driven social media platforms, in different ways, incentivise division, extreme speech and undervalue nuance in storytelling that is needed for people to understand the world. They blend reported and verified reporting with commentary. They encourage that blurring, despite the fact that that in turn is eroding trust in the information environments they are responsible for.

And we see widespread antagonism and hostility towards information producers and information institutions as a result. In an engagement-based paradigm, this matters little. An antagonistic or hostile click is valued the same as a supportive one, but as a measure of the health of the relationships holding up our information ecosystems, it is troubling.

Our research did not find extensive evidence of hate speech that would be easily identifiable. One possible explanation is that this is because the moderation systems designed to identify and report clear examples of hate speech appear to be effective and working. But this does not fix the information environment, which is still polluted by harmful but legal commentary that is degrading public debate. It is also backfiring against news organisations by degrading trust and undermining social media platforms' claims to be forces for good.

Where does this leave policy-makers who are interested in correcting this dysfunctional relationship

between audiences, news organisations and social media platforms? How can public interest news reject the engagement-driven paradigm set by social media?

MOVING TOWARDS A PUBLIC INTEREST NEWS-BASED PARADIGM

Our recommendations are designed to move away from the current engagement-based paradigm which values clicks and attention regardless of the impact of the content that is being created and consumed. We need a new shared understanding of the different values that different kinds of content bring - not judged solely by their commercial worth, but by the common good they contribute to society and democracies.

By defining what constitutes public interest news, and differentiating it from other content on the internet, we can begin to find ways for platforms, media companies and regulators to support information that adds value, rather than contributes to harms.

Below, we set out in more detail recommendations for a pathway forward in which we reach new definitions of public interest news and use that to change the incentives for content production and promotion now and in the coming wave of technological change that generative AI will bring. Demos developed these recommendations and is responsible for them alone. But they were developed after a long engagement process with experts, a stakeholder policy workshop and engagement with representatives from government and regulators, including the UK Government Department for Science Innovation and Technology (DSIT) and Ofcom (the UK media regulator), as well as news media and civil society.

WHAT COUNTS AS PUBLIC INTEREST NEWS?

The use of the term 'public interest news' to identify journalism which may receive special protections or permissions has a long and difficult history.

The Joint Parliamentary Committee on the Draft Online Safety Bill suggested that 'public interest' news would be a more appropriate way to distinguish content that might require extra protections from platform moderation rather than 'journalistic' content or 'content of democratic importance', due to the challenges in defining the latter and precedent in legislation around 'public interest'. This recommendation was designed, as the current media exemption in the Bill is, to protect against overmoderation, but does not place particular duties on platforms to actively design services that would promote and protect this kind of content beyond simple content moderation policies.⁹⁷

Various definitions of 'public interest' already exist: The National Union of Journalists's definition speaks to public safety, exposing corruption and freedom of expression⁹⁸; IPSO's includes many similar themes, including health and safety, disclosing unethical or criminal behaviour, and protecting the public from being misled.⁹⁹

The Cairncross Review indicated that 'public-interest news' is that which is crucial for the functioning of a democracy - with particular emphasis placed on investigative journalism exposing abuses of power, and reporting of the activities of political institutions at a national and local level.¹⁰⁰ The Public Interest Journalism Working Group for the Scottish Government offers a broad definition of public interest news - as journalism which both meets high ethical and quality standards of production and dissemination, and which also has a democratic or educational benefit to citizens.¹⁰¹

The problem with reaching a definition of public interest journalism is that if you ask the politicians, regulators or the platforms to do it, it is rightly seen as an assault on the freedom of the press; if you ask the press to do it, there is a clear interest in defining it as widely as possible in order to support their business models, rather than the quality of the information environment. Instead, we propose that citizens should be engaged in a deliberative process to define what constitutes public interest news, in order to provide a definition that has shared legitimacy that can subsequently be used to create new incentives to produce and promote it.

97 <https://committees.parliament.uk/publications/8206/documents/84092/default/>

98 <https://www.nuj.org.uk/about-us/rules-and-guidance/code-of-conduct/public-interest.html>

99 <https://www.ipso.co.uk/editors-code-of-practice/>

100 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/779882/021919_DCMS_Cairncross_Review_.pdf

101 <https://www.gov.scot/publications/scotlands-news-towards-sustainable-future-public-interest-journalism/pages/3/>

RECOMMENDATION ONE: The government should fund an independent People's Commission on Public Interest News to involve the public in a deliberative way to devise a definition of public interest news

Such definitions have been avoided by the news industry to protect their independence and to prevent compromises to press freedom. But by involving a diverse group of citizens in the process rather than other self interested parties, the process itself would be democratised.

A citizen-led definition of public interest news would provide a definition that could liberate ways to publicly fund news where there are market failures, to incentivise and facilitate production of public interest news by setting up a framework through which platforms would support these efforts through digital competition frameworks. It would also give a definition by which regulators can hold platforms to account for the information they carry and promote.

APPLYING THE PUBLIC INTEREST NEWS DEFINITION

To shift from an engagement-based paradigm to one focused more on public interest news needs different commercial incentives for news companies, different business models and new deals with social media companies. There have been three significant policy interventions relevant in this policy space in recent years: The 2019 independent Cairncross Review into sustainable journalism in the UK;¹⁰² The Online Safety Act;¹⁰³ and the forthcoming Digital Markets, Competition and Consumer Bill, which is expected to include provisions for a news bargaining code, to govern the relationship between news organisations and the platforms and see news organisations more fairly paid for their content.

The Cairncross review made a series of recommendations designed to rebalance the relationship between the news industry and the social media giants in order to fairly compensate them for their content but also to put obligations on the social media companies to improve the content they choose to serve. It also recommended more public funding for high quality journalism via an Institute for Public Interest News. The recommendation for public funding for public interest news was dropped after opposition from the industry to any situation in which the government was setting the terms of what high quality journalism might consist of, citing press freedom.

Efforts to make social media companies fairly pay for the news they serve on their platforms are being taken forward in the forthcoming Digital Markets Bill (which we will come back to).

The 'News Quality Obligation' requiring online platforms to improve their information environments under regulatory supervision argued that: 'Online platforms have already developed initiatives to help users identify reliability, and the trustworthiness of sources. They must continue and expand these efforts, but do so with appropriate oversight. This task is too important to leave entirely to the judgement of commercial entities.' The initial requirement suggested was simply a reporting one, to allow a regulator to gather information on the steps online platforms are taking to improve people's awareness of the origins and quality of the news they read.

Our research has shown how the business models are driving online harms in the form of division and extreme content, while remaining within the law. The examples we saw around conspiracism and the resulting backlash against the media and social media companies suggest that it's not only in society's interest, but also in the news and technology companies' interests to shift from a model that rewards clicks over public interest news.

Our following recommendations are therefore designed to incentivise and sustain news companies to focus on the production of public interest news and for technology companies to fairly promote that news.

RECOMMENDATION TWO: Implement Cairncross by providing public funding for public interest news

There is a market failure in public interest news in a digital world - there are not enough incentives to produce it - particularly at a local level where the dearth of local news is an aggravating factor in community cohesion.

102 <https://www.gov.uk/government/publications/the-cairncross-review-a-sustainable-future-for-journalism>

103 <https://bills.parliament.uk/bills/3137>

In this context, the government should look at providing funding, via an Institute for Public Interest News and using the definition of public interest news created by the People's Commission, to incentivise democratically important journalism.

RECOMMENDATION THREE: The development of digital competition legislation, including how news bargaining codes will work, should set out how the value of public interest news can be recognised by social media platforms, and such content appropriately promoted

The chronic lack of revenue within the journalism sector came up several times across our interviews, as a key driver of the worst outcomes in the information ecosystem. Attention has now turned significantly to how social media platforms which dominate our online environments could be made to financially support healthier environments, and news organisations in particular.

There have been attempts across multiple countries to introduce a news bargaining code, requiring large social media platforms which benefit from news content being shared on their platforms to distribute a fair portion of that value back to news producers.¹⁰⁴ In the UK, this would form part of the Digital Markets regime as a competition measure, and provide a regulatory framework for publishers to negotiate terms with platforms.¹⁰⁵ These regimes in other countries have led to much controversy, as Google and Facebook threatened to leave the Australian market if the news bargaining code led to payment having to be made for news links on their platforms.¹⁰⁶ Similar threats have occurred in Canada in response to the Online News Act, with potentially similar moves in the EU.^{107,108}

Evaluating the impact on citizens' access to information would be a means to ensuring that any proposals which seek to redistribute value from social media companies to media organisations do so in a way that benefits the public good and incentivises better information environments rather than maintaining the status quo.¹⁰⁹

The current proposals are based on paying news producers for sharing links to their content. There are concerns that this does not challenge the problem of reliance on social media engagement for news revenue. Indeed, it may reinforce it, if the path to revenue is greater audience engagement in social media environments.¹¹⁰ Ball argued that these moves could also be ineffective, as news is simply not a large percentage of content shared on platforms. There is a further challenge of defining the news publishers who should be at the table. Attempts to define 'news publishers' in the context of the Online Safety Bill led to widespread criticism that the resulting definition could easily be weaponised by extremists who promote their content as 'journalism', while a narrower definition risks prioritising legacy media over smaller, independent outlets.¹¹¹

There has been some success in Australia, however, outside of the link-sharing model. Although formal designations have not been made of the companies as subject to the code, Google¹¹² and Facebook have entered into agreements with media outlets for compensation, likely to avoid being designated and so eligible for enforced formal arbitration.^{113,114}

What is needed is a way for digital competition regulation to be designed to promote better democratic discourse, rather than focusing narrowly on payment for the most engaged with content or with all news publisher content.

This could look like requiring the negotiations between large social media platforms and news publishers to be based not purely on commercial value. We propose this would look like additional rewards - a higher value - for public interest news produced by news organisations and shared by the social media platforms. This

104 <https://newsmediauk.org/blog/2023/07/04/a-united-uk-news-industry-calls-on-government-to-create-level-playing-field-with-tech-platforms/>

105 https://assets.publishing.service.gov.uk/media/6273af6be90e0746c882c361/Platforms_publishers_advice._A.pdf

106 <https://www.wired.co.uk/article/australia-media-code-facebook-google#:~:text=Google%20and%20Facebook%20did%20not,news%20publishers%20for%20their%20content.>

107 <https://www.nytimes.com/2023/06/23/business/media/meta-google-canada-news-facebook-instagram.html>

108 <https://www.wired.co.uk/article/facebook-is-giving-up-on-news-again>

109 <https://www.smh.com.au/national/a-pox-on-facebook-but-also-on-the-media-bargaining-code-20210218-p573qb.html>

110 <https://www.smh.com.au/national/a-pox-on-facebook-but-also-on-the-media-bargaining-code-20210218-p573qb.html>

111 https://demos.co.uk/wp-content/uploads/2022/05/Exemptions_Exceptions_and_Exclusions__OSB__vF.pdf

112 <https://about.google/google-in-australia/an-open-letter/>

113 <https://www.accc.gov.au/by-industry/digital-platforms-and-services/news-media-bargaining-code/news-media-bargaining-code>

114 <https://www.wired.co.uk/article/australia-media-code-facebook-google#:~:text=Google%20and%20Facebook%20did%20not,news%20publishers%20for%20their%20content>

would shift the incentives to produce more public interest news.

Meanwhile, platforms should also have to demonstrate to regulators how they are promoting public interest news to their audiences. Both these interventions would adopt the public interest news definition from the People's Commission, giving the categorisation legitimacy. This would reduce the risks associated with news bargaining codes while still establishing a pipeline of revenues from social media platforms to public interest news, redistributing power while shifting the incentives to produce news primarily for engagement.

Other ways to support funding of news and platforms' contributions to this could also be explored, separately from discussions around platforms paying for content on their sites. This could look like hypothecated taxes on social media platforms which would be invested in public interest news: redistributing value without the intermediary issue of volume or engagement of content produced. This could be redistributed to news companies based on their production of public interest news (again, defined by the People's Commission).

THE NEXT CHALLENGE: GENERATIVE AI AND THE NEXT STAGE OF DISRUPTION TO THE INFORMATION ENVIRONMENT

The world doesn't stand still. In the course of researching this report, the challenge facing the information environments is poised to shift markedly again. The advent of widely accessible Generative AI is likely to dramatically change the way that news is produced and consumed, bringing a whole new set of challenges for information environments.

Social media is currently the major gateway to news content, particularly for younger audiences,¹¹⁵ but as generative AI tools become more integrated with search and with other online platforms and digital services, it is likely that their use by citizens as sources of information themselves will accelerate.

This brings with it considerable challenges. In the short term, the models are not reliable enough to serve up accurate information.¹¹⁶ At the moment some of the most widely-used models, including GPT-3.5, are not serving up recent information partly to prevent amplification of misinformation. But models such as GPT-4 have improved significantly, and the ambition appears to be for generative AI models to be the dominant way of navigating the world's information before too long. The power and personalisation of AI will shift our information environments again.

Discussants in our policy workshop pointed out that when citizens interact with news information on social media sites, the news sources still retain some power over shaping that interaction. This is markedly different with generative AI tools in which content is produced by the tool, but there is no interaction with the news organisation that may have contributed to the production of that information at all. These tools will also potentially increase the personalisation and individuality of citizens' information environments (as Harari argues) with potential for more individual utility, but may also create more and more narrow echo chambers and further fuel division.

We heard during our workshop that there can be two approaches to the use of generative AI tools: one, to educate the public on the user case and make it clear that they are not tools to access information, although recognising their creative value for other kinds of tasks. The credibility and accuracy of these tools would in this case not be a pressing issue. However, for this to be successful, it would require significant changes in how these tools were described, used by the public, or integrated into existing services such as search - a trend unlikely to reverse without regulatory intervention.

The other direction would be to embrace generative AI's role in facilitating greater access to information in a variety of contexts. This would require significant changes to how these tools are currently being developed. This includes much greater transparency about the design, systems and processes rather than only outputs of large language models. It would include the data on which models are trained, as well as about the process of data labelling and the testing, fine-tuning and accuracy of the resulting models. In particular, the labour behind the labelling of data - much like the current outsourcing of content moderation - is likely to need much more transparency and investment to ensure it is properly resourced, compensated and quality assured.

Discussants in our workshop raised that the generative AI tools which are becoming widely used are not currently of a standard high enough that news organisations feel comfortable allowing their content to be

115 <https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2022>

116 <https://algorithmwatch.org/en/bing-chat-election-2023/>

scraped and used for training the AI models due to the credibility risks it poses. Currently, media organisations such as the BBC and the Guardian have blocked OpenAI from using its content¹¹⁷ over copyright concerns, in the absence of a clear framework through which content could be appropriately licensed and used in AI.¹¹⁸

There will be an increasing demand for high-quality generative AI, which is more reliable and surfaces better quality information than, for instance, the first generation of ChatGPT models which were frequently offering users hallucinations. There is thus a role for high-quality, independent journalism, providing content which has been produced robustly and ethically to enable issues of generative AI tool accuracy to be at least partially mitigated.

RECOMMENDATION FOUR: Media organisations should develop a standards code for companies producing generative AI tools to use their content

This would set out the conditions of design, development and deployment of these tools under which the media organisation would permit its content to be licensed to train the tool. It would set standards for accuracy and sourcing of information. These codes should be developed in consultation with civil society and where appropriate, regulators.

It is our view that it is sensible for news organisations to restrict the use of their data in LLM training sets while this is developed in order to protect their commercial value and mitigate against any reputational risks of generative AI misusing their content.

RECOMMENDATION FIVE: The government should include a news bargaining framework for AI platforms in the forthcoming Digital Markets Bill to ensure that where news is being used in LLMs, news organisations are fairly compensated

Again, this should value public interest news more highly than other content, using the new definition of public interest news created by the citizens-led process.

These recommendations are designed to ensure that the early development of generative AI models understands the value of public interest news from the outset and avoids a situation whereby regulators are trying to retrospectively claw back compensation for the industry in the future. By elevating public interest news from the outset, there would be a better chance of maintaining a healthy information environment in the long term.

The first four recommendations are designed to define public interest news and recognise its value accordingly via public funding, new settlements with social media companies and future settlements with generative AI companies. We now set out three further recommendations that flow from the research and insights in this report.

1. THE ROLE OF CITIZENS

There is one clear message from this research: the social media platforms are responsible for much of the dysfunction in our information environments, but news media organisations and audiences also have a proportional role to play in improving online discourse.

As Purpose Australia have reported in similar work examining the relationship between news, social media and audiences: "Responsibility for detecting and countering hate speech should not sit with the public, but there is a need to improve communities' and individuals' media and digital literacy - as this enables their full participation in society. The scale of this requires investment by governments and industry, and should include building awareness of relevant legislation and industry codes of practice. This will better equip communities and individuals to understand, identify and report dehumanising hate speech – as well as knowing their rights in relation to the issue."¹¹⁹

We wholeheartedly agree with this assessment.

117 <https://www.telegraph.co.uk/business/2023/10/06/bbc-blocks-chatgpt-content-use-ai-copyright-concerns/>

118 We welcome positive steps in this direction, such as: <https://rsf.org/en/rsf-and-16-partners-unveil-paris-charter-ai-and-journalism>

119 <https://www.purpose.com/australian-news-media-and-online-hate-speech/#:~:text=Online%20hate%20speech%20is%20a,%2Djerk%2C%20counterproductive%20policy%20design; Interviewee13>

RECOMMENDATION SIX: The government should fund a major digital literacy programme across educational institutions but also pilot ways to educate other online audiences

2. LIBERATING RESEARCH INTO INFORMATION ENVIRONMENTS

There is a worrying trend towards social media companies restricting access to data for researchers to study online environments. Researchers are also facing lawsuits from tech companies for calling out online harms.¹²⁰ Policymakers are taking some steps in this direction by empowering regulators to have greater transparency and information powers, but there is an urgent need to allow civil society and academia to scrutinise this data and conduct research projects like this one.¹²¹

RECOMMENDATION SEVEN: The UK should follow the EU's example and legislate to secure data access from major online platforms for public interest research.^{122,123}

International governments, including the US, should also work together to secure a consistent international programme around this issue. Our information environments are our spaces of public discourse and democratic debate. They shouldn't be controlled by private actors - we need to understand the world we operate in.

3. NEXT STEPS ON ONLINE HARMS

In the UK the Online Harms Act has just been passed in Parliament. The major attempt to improve our online environments is now law. We have written elsewhere¹²⁴ about the missed opportunity this presented to properly tackle the wider issues of legal but harmful content online. An obligation for platforms to be accountable for 'legal but harmful' content was dropped after a political backlash protecting freedom of speech and about what constituted an online harm. The final Bill, as it is passed in the Commons, holds platforms to account for their moderation of illegal content, enforcement of terms and conditions, and protection of children from accessing harmful content.

Ofcom will have a duty under the Online Safety Act to form an advisory committee on disinformation and misinformation, and advise on steps which social media platforms could take to reduce the risks of these information harms on their services. This should take into account the diverse forms that 'disinformation' can take, as demonstrated in this report - from amplification of conspiracy theories to amplification of harmful narratives that target particular marginalised groups.

RECOMMENDATION EIGHT: Ofcom's advisory committee on disinformation and misinformation should take a wide view of information harms

120 <https://www.theguardian.com/technology/2023/aug/08/techscape-elon-musk-x-twitter-lawsuit-ccd-h-hate-speech>

121 https://www.reset.tech/documents/Civil_Society_Briefing_OSB_Data_Access.pdf

122 <https://algorithmwatch.org/en/dsa-data-access-explained/>

123 <https://algorithmwatch.org/en/call-for-evidence-data-access-platform-researchers/>

124 <https://www.politicshome.com/thehouse/article/we-can-no-longer-support-the-online-safety-bill>

CORE RECOMMENDATIONS

WHO	WHAT	WHY
Government	The government should fund an independent People's Commission on Public Interest News to involve the public in a deliberative way to devise a definition of public interest news	To establish a democratically legitimate definition of public interest news that all parties can buy into, in order to reset the value of different types of content in the information environment
Government	Implement Cairncross by providing public funding for public interest news	To address market failures in the production of public interest news
Government	The development of digital competition legislation, including how news bargaining codes will work, should set out how the value of public interest news can be recognised by social media platforms, and such content appropriately promoted	To apply an appropriate value to public interest news not solely based on engagement, and to hold platforms to account for appropriately promoting public interest news
Media companies	Media organisations should develop a standards code for companies producing generative AI tools to use their content	To ensure that public interest news is protected and valued in future generative AI systems
Government	The government should include a news bargaining framework for AI platforms in the forthcoming Digital Markets Bill to ensure that where news is being used in LLMs, news organisations are fairly compensated	To ensure that public interest news is protected and valued in future generative AI systems from the outset

SUPPLEMENTARY RECOMMENDATIONS

WHO	WHAT	WHY
Government	The government should fund a major digital literacy programme across educational institutions, but also pilot ways to educate other online audiences	To empower citizens to better assess the information they see online
Regulators	The UK should follow the EU's example and legislate to secure data access from major online platforms for public interest research ^{124,125}	To ensure our public information environments are understood and scrutinised, providing a sound knowledge base for policy makers to work from
Regulators	Ofcom's advisory committee on disinformation and misinformation should take a wide view of information harms to ensure it captures all forms 'disinformation' can take	To encourage platforms to improve social media environments to reduce the risks of information harms on their services

125 <https://algorithmwatch.org/en/dsa-data-access-explained/>

126 <https://algorithmwatch.org/en/call-for-evidence-data-access-platform-researchers/>

CONCLUSION

In this report we sought to understand how digitalisation has changed news media and journalism, how information producers and consumers are interacting in social media spaces, and what ramifications this has for the health of digital discourse.

Digital ecosystems are enabling new freedoms of expression but also new forms of suppression, while digitalisation encourages pluralism in theory but in practice can bake in incumbency, both in the case of the big tech platforms and the legacy media. And we found that although information environments should help bridge social divides at the local level, too often they are deepening existing rifts.

Looking more closely at the specific interactions between news organisations and citizens in online spaces through three highly engaged-with and controversial news stories this year, we found a similar story. Audiences are key drivers of digital discourse, but this operates within a social media engagement-based paradigm which does not lend itself to meaningful participation, but instead a vicious cycle of controversy and backlash feeding more controversy. Democratic discussion is intertwined with division, hate, and false or misleading narratives, allowing them to be amplified. As a result, information chaos persists.

People are not satisfied with the quality of democratic discourse or information production, and distrust in media institutions is endemic.

If this is to be repaired, we need to focus on how to rebuild the relationship itself, not only how to tinker with the products of information which are produced or consumed. We need a new shared understanding of public interest news, to help challenge the engagement-based paradigm and shift towards a public interest-based paradigm that promotes social good and democratic discourse.

We believe that to achieve this requires a fundamental shift in the relationship between citizens and news information, which can be achieved through citizens being involved in defining public interest news, and that definition being used to reset the terms of engagement between governments, news producers and social media platforms, and to reward democratically important information.

Digital technologies offer unprecedented opportunities for citizens to exercise their democratic and human rights. But without a meaningful redistribution of power, we will continue to fall into the same patterns of disconnection and division, with marginalised communities suffering the most.

With the rapid development of AI, we are on the brink of a new era of digital policymaking. Now is the time to break the cycle, and change the structures of our information ecosystems to put people back at the heart of public interest news.

ANNEX ONE

METHODS

SECTION 1: SUMMARY

Our initial dataset included 196,448 Facebook posts, 16,039 Instagram posts and 12,453 YouTube videos posted by an account belonging to an influential news organisation in the UK or US between 01/03/2023 and 13/07/2023.

Selection of outlets

We wanted to begin with mainstream media companies in the UK and the US which have substantial online influence in general, not just on social media. We used Ipsos data on UK audience numbers for news websites¹²⁷ and SimilarWeb data on US news website visits as proxies for online influence, and listed the 15 top UK and top 15 US news websites.¹²⁸

We then inspected these lists and manually added news websites which, while not in these lists, had appeared in interviews or desk research as worth particular consideration - either due to their influence with elite decisionmakers, and/or as particularly interesting emergent players. These were The Financial Times, GBNews, and TalkTV.

For each media outlet we found their Facebook, Instagram, Twitter (now X), and YouTube accounts, and selected the top 10 for each of the UK and US by total number of followers.¹²⁹ We had intended to manually ensure a variety of ideological leanings and types (to ensure inclusion of, for example, tabloids, 'high status' papers, online-first news, news-focused vs. broader content). However we judged that a simple ranking provided sufficient variety without needing to manually adjust.

TABLE 1 (FROM PART ONE)

SELECTED INFLUENTIAL NEWS ORGANISATIONS FOR COLLECTION OF SOCIAL MEDIA POSTS

UK NEWS ORGANISATIONS		US NEWS ORGANISATIONS	
• BBC News	• ITV	• ABC	• NBC
• Daily Mail	• The Mirror	• BuzzFeed	• The New York Times
• The Financial Times	• Sky News	• CNN	• People
• The Guardian	• The Sun	• Forbes	• The Washington Post
• The Independent	• The Telegraph	• Fox News	• The Wall Street Journal

127 https://pressgazette.co.uk/media-audience-and-business-data/media_metrics/most-popular-websites-news-uk-monthly-2/; <https://www.similarweb.com/top-websites/united-kingdom/news-and-media/>; https://en.wikipedia.org/wiki/Mass_media_in_the_United_Kingdom

128 https://pressgazette.co.uk/media-audience-and-business-data/media_metrics/most-popular-websites-news-us-monthly-3/; <https://www.similarweb.com/de/top-websites/united-states/news-and-media/>; <https://www.pewresearch.org/short-reads/2021/05/07/broad-agreement-in-u-s-even-among-partisans-on-which-news-outlets-are-part-of-the-mainstream-media/>

129 Where they had multiple accounts per platform, we only selected the account with the highest number of followers. We also made separate lists which also included all the accounts for each platform, plus any LinkedIn, and TikTok accounts; however the top 10 were still the same.

Platform affordances play a key role in how news is shared and engaged with: for instance, Instagram’s limited link-sharing functionality means that it is more difficult for outlets to link to published articles off-platform, while YouTube is easier for broadcast media to adapt to than for other traditional media, but video content generally takes more time to produce than written content. These differences influence how much different sources post on different platforms, and the type of content they upload. X (formally Twitter), although widely used by news organisations, was out of the scope of this research due to data access restrictions introduced by Elon Musk in 2023.

Within the whole dataset, several different kinds of news and media content were identified. 358 topics were produced by topic modelling - a Natural Language Processing method, outlined below - Facebook/Instagram posts and Youtube video titles. We identified multiple key themes, reviewing the most posted-about topics and the most commented-on topics. These themes often overlapped, so were coded according to the most relevant category (for instance, stories about a famous political person such as Donald Trump were coded as ‘politics’ due to the strong political context and relevance of the stories).

TABLE 8
QUALITATIVE CODING OF TOPICS PRODUCED BY TOPIC MODEL

THEME	EXAMPLE HEADLINES
Politics: Political news, including international, domestic, or geopolitical	House approves debt ceiling deal. Here’s how representatives voted.
Lifestyle: Focusing on health, money, advice, personal life	‘They Don’t Like My Husband. But We Can Still Be Friends, Right?’
Pop culture: Discussion of TV, film, music, sports	Eurovision Song Contest 2023: When, how to watch and who to look out for
Famous People: Stories which focus on individual famous people	Bob Saget’s widow Kelly Rizzo shares how to best honor the ‘Full House’ star on his birthday
Disaster: deliberate or accidental human disasters such as shootings or attacks, natural disasters: where focus of story is the event and the human impact	Titanic sub live updates: Safety investigations launched into Titan implosion deaths
Judicial: Discussions of trials, courts, legal proceedings: including criminal and civil	Alec Baldwin: All criminal charges against actor over fatal shooting on Rust set are dropped
General Public Interest: Discussion of socioeconomic trends affecting people’s lives - public services, work, money	Patients in England to be asked to use NHS app to book private hospitals
Identity: Where the main focus of story is on an identity-based issue	Riley Gaines slams Karine Jean-Pierre’s response to question about trans participation in girls sports
Media: Stories focusing on media organisations or media personalities	Tucker Carlson leaves Fox News
Other: e.g. ‘feel-good’ stories about local events	Firefighters rescue puppy after getting her head stuck in a wall

From the topic model we were able to identify which news outlets were the most active, what platforms they engaged with the most, and across what topics they provided the most coverage.

TABLE 9

TOP 3 NEWS ORGANISATIONS BY NUMBER OF POSTS, PER PLATFORM

PLATFORM	NEWS OULET	PLATFORM COUNT	TOTAL COUNT	MOST COVERED TOPICS ON PLATFORM
Facebook	Daily Mail	18692	19657	Titanic submersible, Emily Morgan, Bud Light backlash
	The Sun	14678	19657	DIY, Home renovation, Shopping
	The Telegraph	14624	16048	King Charles III coronation, Matt Hancock lockdown, Prigozhin/Wagner group
Instagram	Fox News	3086	10609	Riley Gaines / Transgender athletes, Bud Light backlash, Hunter Biden whistleblower
	BuzzFeed News	2613	14071	Relationships, Movies, Actors/Actresses
	People Magazine	1484	13176	Titanic submersible, Met Gala, Dogs
YouTube	Forbes	954	15767	Chris Wray / FBI weaponization, Marjorie Taylor Greene, Chip Roy
	The Independent	951	11889	Rishi Sunak, David Cameron Covid inquiry, Prigozhin/Wagner group
	Fox News	907	10609	Greg Gutfeld, Jeanine Pirro, Hunter Biden Whistleblower

Of these, the most relevant for our purposes were those relating to democratic debates, live political issues, discussion of the information ecosystem itself, and where there was the potential for discourse to tip into online harm. As such, stories which were relating to politics, media or identity were classified as relevant, using a human classification process outlined below. This gave us a total of 98 topics that were relevant, and 250 irrelevant.

We then examined the most posted about and most engaged with topics, looking across numbers of articles/videos posted across platforms within each topic, and then the number of replies the stories in each topic had received to select case studies for further investigation. Due to data access restrictions, replies from Facebook and Instagram were not directly available for collection. Instead we needed to extract URLs from the collected posts and perform platform-wide searches for mentions of these URLs.

The top 11 included:

- Bud Light's partnership with Dylan Mulvaney
- Phillip Schofield, Holly Willoughby and This Morning (e.g. the controversy around Schofield's departure from the programme and relationship with a colleague)
- Hunter Biden and the FBI (e.g. the investigations into allegations of corruption or criminal wrongdoing by Hunter Biden)
- Gary Lineker and Match of Day (e.g. Lineker's comments about the Government's refugee policy and his subsequent suspension)
- Tucker Carlson and Fox News (e.g. Tucker Carlson being fired from Fox)
- The case of a subway rider, Jordan Neely, in New York who died after being subjected to a chokehold by a fellow passenger¹³⁰
- The Manhattan District Attorney Alvin Bragg (who announced Trump's indictment)¹³¹
- The defamation lawsuit brought by Dominion against Fox News¹³²
- Drag performances in Tennessee (e.g. when public performances were banned)¹³³
- The controversy around Huw Edwards, the BBC presenter, after allegations emerged about him paying a younger person for sexual pictures¹³⁴
- Robert F Kennedy (who had been making a bid to be the Democratic candidate in the US 2024 election)

Of these, we selected Bud Light, Hunter Biden and This Morning as the stories with high engagement levels across the different measures of engagement, as well as representing stories both across the US and UK divide.

130 <https://www.bbc.co.uk/news/world-us-canada-65578905>

131 <https://manhattanda.org/district-attorney-bragg-announces-34-count-felony-indictment-of-former-president-donald-j-trump/>

132 <https://www.theguardian.com/us-news/2023/apr/18/fox-dominion-settle-us-defamation-lawsuit>

133 <https://news.sky.com/story/tennessee-becomes-first-us-state-to-ban-public-drag-performances-12824159>

134 <https://www.bbc.co.uk/news/entertainment-arts-66186358>

TABLE 10

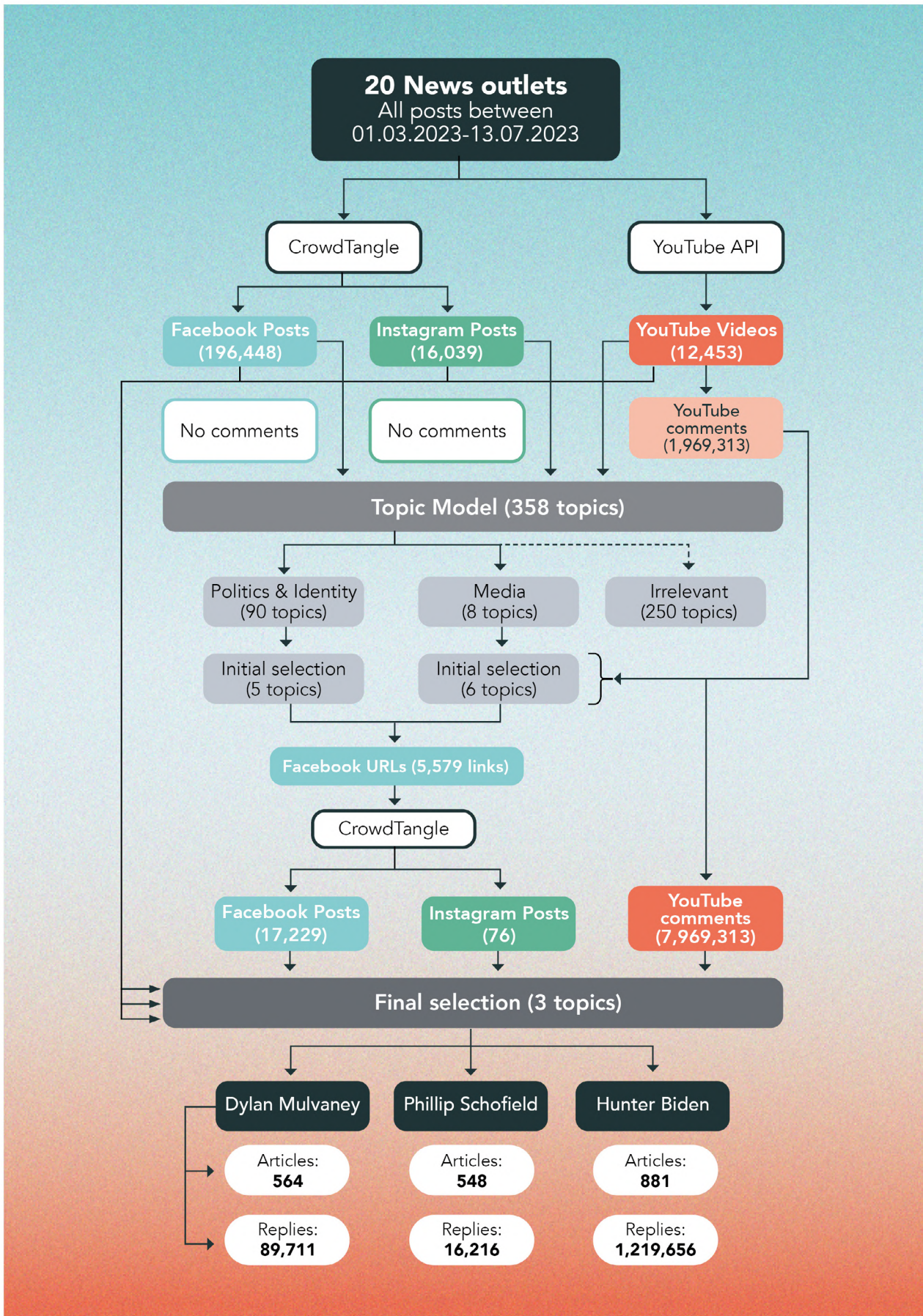
LEVELS OF ENGAGEMENT (NUMBER OF POSTS, REPLIES AND LIKES) FOR EACH OF OUR 3 SELECTED CASE STUDIES, ACROSS ALL 3 PLATFORMS

	FACEBOOK COUNT	INSTAGRAM COUNT	YOUTUBE COUNT	TOTAL COUNT	FACEBOOK REPLIES	YOUTUBE REPLIES	TOTAL REPLIES	FACEBOOK LIKES	YOUTUBE LIKES	TOTAL LIKES
Bud Light	2938	9	27877	30824	240563	12797	253360	272700	120015	392715
Hunter Biden	5185	37	953507	958729	229615	350412	580027	280853	3429531	3710366
This Morning	1661	0	20458	22119	142203	8316	150519	119744	113738	233482

Amongst the posts which contained the hyperlinks to articles of interest from Facebook and Instagram, our hypothesis was that there would be high levels of engagement not only with media outlets sharing their own stories, but also with media personalities discussing and sharing their own takes on news stories. However: this was not the case as the majority of the amplification of these stories was being done by other media outlets that were, most of the time, sub-divisions of the originating actor organisation.

METHODS

FIGURE 1 (FROM SECTION ONE)
DATA ANALYSIS PIPELINE



DATA COLLECTION

The methodology began with data collection from Facebook, Instagram and Youtube, between the beginning of March (01/03/2023) and the middle of July (13/07/2023). The type of data collected was broken down into two categories, posts and replies. For Facebook and Instagram, posts were collected from pages owned by selected media sources (the 'Actor list' below) promoting stories and linking to their external websites. For YouTube, videos from profiles in the actor list were collected. The replies cover the direct responses to these posts, which for Facebook and Instagram was done through link analysis (finding other posts that contained links shared by pages in the actor list) , and for YouTube the comments responding to each video were collected.

Data was collected from all 3 platforms using the list of 20 actors provided below, 10 US-based news organisations and 10 UK-based. The starting collection aimed to collect all posts within the mentioned time frame, covering all of their journalist activity across the 3 platforms. For Facebook and Instagram we utilised the CrowdTangle API, and for YouTube the YouTube Data API.

Actor list:

'ABC News', 'BBC News', 'BuzzFeed News', 'CNN', 'Daily Mail', 'Financial Times', 'Forbes', 'Fox News', 'ITV News', 'NBC News', 'People Magazine', 'Sky News', 'The Guardian', 'The Independent', 'The Mirror', 'The New York Times', 'The Sun', 'The Telegraph', 'The Wall Street Journal', 'The Washington Post'

YouTube replies were collected immediately, following the completion of the post collection, with an additional 2 days added to the end date (15/03/2023) to allow for suitable representation of stories captured towards the end of the collection period.

Compliance

All data was collected in agreement with the respective terms and conditions of the relevant platforms and their API's. Data handling, anonymisation, sharing and privacy was covered and approved in a submitted ethical compliance application to The University of Sussex Cross-Schools Research Ethics Committee.

Initial data counts

The below table contain the volume of data collected for each platform in this first round of collection:

TABLE 11
COUNTS OF POSTS AND REPLIES INITIALLY COLLECTED FROM EACH PLATFORM

PLATFORM	POSTS	REPLIES
Facebook	196,448	N/A
Instagram	16,039	N/A
YouTube	12,453	7,969,313

TOPIC MODELLING

With the collected posts, topic modelling was deployed to group narratives and stories into digestible topics which could then further be analysed. Topic modelling is an unsupervised machine learning technique that identifies semantic similarities in a document space and groups them into clusters - in the case of this project, finding headlines which use similar language and/or refer to similar ideas and thereby draw out recurring topics in the headlines. In order to create a uniform input to the topic model, the text of Facebook and Instagram posts were added to the titles of the attached articles, so that when combined with the YouTube data there would be a suitable amount of text amounting to a 'headline' for the topic modelling.

With ample tuning of various parameters and observing changes in outputs, posts within the same cluster are expected to be covering the same/similar stories with a minimal amount of outliers. Our initial model with un-tuned hyperparameters led to the generation of 3,907 distinct topics. We found that this model was too granular, causing key figures such as Donald Trump, who received a lot of coverage over several different but linked stories, to be the centre of a large number of topics, which ideally should have been grouped together as one. We improved upon this model by tuning the minimum size of a cluster to 100. This greatly reduced the produced topics down to 358 which analysts then manually labelled.

Theme identification

The next step was to narrow down the list of candidate topics to ones that would make an interesting and relevant investigation. Although most posts were related to current affairs in politics and the media, there was a significant amount that were not relevant to the investigation, such as recipes, or fitness.

During the labelling of the topic model, analysts identified 4 different top-level themes that could be used to categorise topics. These were: Politics, Media, Identity, and Irrelevant (containing everything not in the previous 3 themes). The following table shows the number of topics that fell into each theme:

TABLE 12
INITIAL QUALITATIVE CODING OF TOPICS

LABEL (THEME)	COUNT
Irrelevant	240
Politics	81
Media	8
Identity	29

Theme evaluation

In order to assess the performance of our topic model, and theme allocation, a sample of 1,000 posts was randomly selected and labelled according to which of the above themes each post belonged to. Each post had an annotation for the topic it was assigned to, and thus a predicted theme according to the topic-theme mapping that was created. We evaluated the performance by comparing the assigned theme from the topic model to the manually annotated theme, with the initial results as follows:

TABLE 13
PERFORMANCE METRICS OF INITIAL TOPICS

THEME	PRECISION	RECALL	F1 SCORE	CLASS SIZE
Irrelevant	0.76	0.95	0.84	664
Politics	0.76	0.48	0.59	189
Media	0.79	0.50	0.61	22
Identity	0.55	0.17	0.26	125

With the F1 Score, a measure of the overall accuracy of a model on a particular dataset, being our main metric of performance, we were not satisfied with the initial results and performed a qualitative analysis of the results, looking at where misclassifications were occurring. To summarise we found the following:

- Politics and Identity had a strong thematic intersection. Posts labelled as Identity regularly fell into topics labelled as Politics, and vice versa.
- A number of topics that had been labelled as one of the relevant themes did not agree with the labelling of the individual posts
- A large number of relevant posts fell into the -1 outlier or 'irrelevant' topic, i.e. they could not be easily assigned to any of the existing topics and there were not enough of them to form a new topic (i.e. they did not form a cluster larger than the minimum size specified in the model hyperparameters, 100). As a result they received the -1 topic label, which automatically classified them as Irrelevant.

For each of the above issues we came up with a suitable solution to improve performance, as described below.

Politics and Identity intersection

To solve the issue of these two themes intersecting, analysts conducted a re-labelling of all the posts that were labelled as Identity. Because the model clusters political issues more strongly than identity issues, we decided that any post discussing policies and legislation that affect people's identity would be labelled as Politics.

Upon re-evaluating the data, we discovered that after re-labelling, most of the posts relating to Identity had been changed to Politics, leaving very few 'Identity' posts to evaluate on. Although we managed to improve the F1 score of this theme slightly, the lack of evaluation data along with the similarity in the themes led to the decision of combining Politics and Identity into a single theme.

Irrelevant topics

Upon inspection it was found that 12 topics labelled as relevant contained a majority of irrelevant documents (according to the evaluation data). These topics were selected for inspection which included more sampling of the documents they contained. The result of this was 10 of the 12 topics were relabelled to being Irrelevant. This in turn reduced the number of cases where irrelevant documents were labelled as relevant by the topic model.

Outlier reduction

The final improvement to the model came from the reduction of outliers. This process involved taking all the posts that had been identified as an outlier (clustered under the -1 topic) and assigning them to the closest non-outlier topic. Using several metrics, we found that the most effective method was to compute the sentence embedding representations for the outlier posts, and join them with the topic centroid that was closest in terms of cosine similarity. The final topic-theme counts and evaluation scores can be found below respectively:

TABLE 14
FINAL COUNTS OF TOPICS

LABEL (THEME)	COUNT
Irrelevant	250
Politics and Identity	90
Media	8

TABLE 15
PERFORMANCE METRICS OF FINAL TOPICS

THEME	PRECISION	RECALL	F1 SCORE	CLASS SIZE
Irrelevant	0.93	0.90	0.92	746
Politics and Identity	0.73	0.81	0.77	232
Media	0.70	0.64	0.67	22

TOPIC SELECTION AND REPLY COLLECTION

With the allocation of the topic themes, a select number of candidate topics were identified for further data collection:

- 5 topics from 'Politics and Identity'
- 6 topics from 'Media'

Reply data was already collected from YouTube, in the form of YouTube comments, but Facebook and Instagram do not allow for the collection of public comments through any of their available channels. Therefore for these two platforms we used the below approach.

With the only content accessible to us being public pages, we decided to utilise the CrowdTangle link searcher to perform a collection using the external article links collected from our actors. For each post belonging to the 11 topics that were selected, the links were extracted, normalised, and expanded and used to search across the Facebook and Instagram platforms. Any post where a match was found was collected and treated as a reply to the original post it was collected from.

Using the CrowdTangle link searcher we collected the following:

TABLE 16
REPLIES TO POSTS FROM ORIGINAL DATASET COLLECTED USING CROWDTANGLE

PLATFORM	REPLIES
Facebook	17,229
Instagram	76
YouTube (from prior collection)	1,615,951

The above counts are from **after** several stages of removals:

- Removing duplicate posts which were in the original post collection
- Removing duplicate posts which were collected twice with different URLs (same CrowdTangle ID)
- Removing posts where the contents contained no message (these were posts that were sharing an article but not responding in any way)

We acknowledge that there is a large discrepancy between the platforms. Instagram does not allow user posts to link to external platforms, which acts as a deterrent for posting links in posts. Comparing Facebook and Instagram to YouTube also shows a lack of completeness. With the restriction of not being able to collect Facebook comments, there is a loss of great amounts of relevant data that could be used to support the investigation.

TOPIC CASE STUDIES

With the collected data, analysis was done into the engagement of different topics, and overall volume, in order to select a series of case studies for further analysis. The analysts ultimately decided on 3 case studies for further quantitative and qualitative analysis.

- Dylan Mulvaney Bud Light controversy
- Phillip Schofield This Morning
- Hunter Biden and related investigations

Each of these topics then went through several stage of analysis to better understand the relationship between information providers and information consumers. Our method was chosen to produce a healthy balance between maintaining a feasible timescale for analysis, while maximising completeness when analysing the dataset.

Post filtering

The first step was to filter out any remaining irrelevancies in the dataset (in the collected posts). One of the faults with current topic modelling approaches is a tendency to have data within the dataset that was not relevant to the current topic. To fix this, keyword filtering was performed to reduce the dataset and increase the precision as much as possible without impacting recall. Keywords were produced through a mixture of manual inspection of relevant articles and keyword extraction techniques. After this the datasets were small enough they could be manually reviewed and labelled. This manual labelling should result in a 1.0 score for the precision of the topics. However, some data may have been excluded in the filtering stages, we cannot therefore claim to have 1.0 recall. On top of this there is also the caveat of potential human error when labelling, which would have a very minor impact on the precision.

TABLE 17
PERFORMANCE METRICS OF TOPICS SELECTED FOR CASE STUDIES

TOPIC	FILTERING PERFORMANCE			COUNTS	
	Precision	Recall	F1	Post-filter count	Post-review count
Dylan Mulvaney	0.97	0.98	0.98	591	564
Phillip Schofield	Manually reviewed ¹³⁴			1,114	548
Hunter Biden	0.79	0.90	0.84	930	811

¹³⁵ For Phillip Schofield, we manually annotated the 1114 articles according to their theme (Affair, General This Morning, Irrelevant) and then selected those under 'Affair' (548).

Post framing coding

During the process of reviewing the remaining posts, an additional series of annotations were coded, according to the framing of the news.

Reply topic modelling

For each story, we had a collection of replies which were used to produce new topic models representing the different types of narratives appearing as a response to the news stories. Due to the unsupervised nature of topic modelling, the volumes of data being highlighted by the model are not totals, but instead a representation of the proportion of replies across narratives.

For each topic produced by the topic model, analysts gave a rough summarisation of the topic with support from a random sample of 100 replies for each topic. In a second round of annotation, those topics were grouped by overarching themes (super-topics). As a result, each of the replies were categorised under an overarching super-topic. Super-topics were specific to the topic of the source news story.

One key observation about the super-topics is that they were not defined by their sentiment. For example: replies that fell into the super-topic about 'Transphobia' for Dylan Mulvaney were not inherently transphobic, but were highlighted as being relevant to discussion around transphobia, including replies being transphobic.

The volumes of super-topic membership for the replies, along with information about post counts, volumes per news outlet and news framing, were all used in order to produce the analysis for this investigation.

Qualitative coding

Throughout this investigation analysts continuously developed qualitative coding frameworks for news organisation posts, and for the topics within the replies which were identified by the topic model. This was an iterative process, involving multiple analysts. Analysts reviewed the data to identify common narratives, allowing the themes to emerge from the data rather than trying to fit the data into predetermined categories. These categorisations were regularly discussed amongst the analyst team to test and refine them; multiple coding was also used to identify and resolve differences in coding. This process also drew on analysts' expertise and knowledge of harmful narratives online to identify topics or keywords of particular interest to the investigation for further analysis.

Methodological investigations: challenges of automated classification

One method that was trialled and rejected early on was a 'sentiment classifier'. This was rejected for a number of reasons including the fact that sentiment isn't always aimed towards the target, for example 'I am so happy that Bud is losing money' expresses a positive sentiment overall, but a negative sentiment towards Bud Light. Determining the target of a particular sentiment became challenging to determine in YouTube comments, as many commenters would reply to each other, rather than directly referring to our relevant actors such as Bud Light. Another challenge was that most news headlines contain neither positive or negative sentiment. As a result, our sentiment classifiers performed poorly.

A hate speech classifier was also considered to identify the most harmful forms of speech. However, hate speech classifiers have variable performance based on the dataset and the target - they also tend to miss out on a lot of subtle hate like deliberate misgendering of trans people. Moreover, platform moderation¹³⁶ made it unlikely that significant amounts of extreme hate speech would be present in the dataset, and we were interested to explore the grey areas of more subtle forms of harm.

We also considered classifying using generative AI. This is an approach where models similar to GPT-4, underlying the popular ChatGPT, are queried with examples of text and are asked to predict which of the provided categories it belongs to. However, the rate limits on most of these models make this impractical to process large datasets, and the reliability can be variable. The methods we deployed likely give a more accurate result, along with being precisely tailored to the definitions we were interested in. This is also a grey area in regard to ethics, as this would involve providing collected personal data to companies which sometimes do not disclose whether or not that data will be used for further training. The methods used for

136 <https://www.theverge.com/2020/12/3/22150197/youtube-comments-posting-hurtful-hate-videos-discrimination-monetization-search>

this investigation were carried out offline - no personal data was run through any model hosted by anyone other than the investigators.

Zero-shot classifiers were trialled for a number of different filtering tasks. These are models that are generally a lot smaller than those used in generative AI, and can be run locally without concerns of where the data is being sent. These models can be provided with text and a list of possible classification categories, and without any prior training can be used to predict which category it belongs to. The effectiveness of these models depends on the domain they were initially trained on. It was found that they were not suitable for our task due to the specific definitions we were working with.

Finally, we investigated the viability of transformer-based classifiers. This involves taking a model that has been pre-trained on a previous dataset (usually a very resource intensive task taking several weeks to months), and then fine-tuning it specifically for a new task. It retains the model of language that it previously developed, and is given examples of new language that it needs to adapt to. In our use case this would have involved feeding the model examples of different types of language we wanted to classify, such as the different supertopics identified in the analysis, and then asking it to predict that class. This method unfortunately is limited by the amount of data that can be provided. Because we have such a diverse range of themes that need to be identified, and several different contexts it needs to take into account, the amount of data required to achieve an acceptable level of accuracy would have been far beyond what was feasible within this project's timeframe.

The challenges in developing automated classification methods to analyse large-scale data of this nature highlight a very real policy problem as well as research challenge. Common propositions to deal with the problem of harmful speech online include wide scale automated moderation and/or curation (e.g. labelling or downranking content). However, when narratives are closely interrelated, as demonstrated here, it will always be a challenge to accurately detect harmful narratives, and to subsequently moderate content in ways that do not also impinge on democratic debate. There is a need to change the terms of engagement, rather than simply respond to the harms arising, which sits behind our call for a more participatory media environment.

Licence to publish

Demos – Licence to Publish

The work (as defined below) is provided under the terms of this licence ('licence'). The work is protected by copyright and/or other applicable law. Any use of the work other than as authorized under this licence is prohibited. By exercising any rights to the work provided here, you accept and agree to be bound by the terms of this licence. Demos grants you the rights contained here in consideration of your acceptance of such terms and conditions.

1 Definitions

a 'Collective Work' means a work, such as a periodical issue, anthology or encyclopedia, in which the Work in its entirety in unmodified form, along with a number of other contributions, constituting separate and independent works in themselves, are assembled into a collective whole. A work that constitutes a Collective Work will not be considered a Derivative Work (as defined below) for the purposes of this Licence.

b 'Derivative Work' means a work based upon the Work or upon the Work and other pre-existing works, such as a musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which the Work may be recast, transformed, or adapted, except that a work that constitutes a Collective Work or a translation from English into another language will not be considered a Derivative Work for the purpose of this Licence.

c 'Licensor' means the individual or entity that offers the Work under the terms of this Licence.

d 'Original Author' means the individual or entity who created the Work.

e 'Work' means the copyrightable work of authorship offered under the terms of this Licence.

f 'You' means an individual or entity exercising rights under this Licence who has not previously violated the terms of this Licence with respect to the Work, or who has received express permission from Demos to exercise rights under this Licence despite a previous violation.

2 Fair Use Rights

Nothing in this licence is intended to reduce, limit, or restrict any rights arising from fair use, first sale or other limitations on the exclusive rights of the copyright owner under copyright law or other applicable laws.

3 Licence Grant

Subject to the terms and conditions of this Licence, Licensor hereby grants You a worldwide, royalty-free, non-exclusive, perpetual (for the duration of the applicable copyright) licence to exercise the rights in the Work as stated below:

a to reproduce the Work, to incorporate the Work into one or more Collective Works, and to reproduce the Work as incorporated in the Collective Works;

b to distribute copies or phono-records of, display publicly, perform publicly, and perform publicly by means of a digital audio transmission the Work including as incorporated in Collective Works; The above rights may be exercised in all media and formats whether now known or hereafter devised. The above rights include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. All rights not expressly granted by Licensor are hereby reserved.

4 Restrictions

The licence granted in Section 3 above is expressly made subject to and limited by the following restrictions:

a You may distribute, publicly display, publicly perform, or publicly digitally perform the Work only under the terms of this Licence, and You must include a copy of, or the Uniform Resource Identifier for, this Licence with every copy or phono-record of the Work You distribute, publicly display, publicly perform, or publicly digitally perform. You may not offer or impose any terms on the Work that alter or restrict the terms of this Licence or the recipients' exercise of the rights granted hereunder. You may not sublicense the Work. You must keep intact all notices that refer to this Licence and to the disclaimer of warranties. You may not distribute, publicly display, publicly perform, or publicly digitally perform the Work with any technological measures that control access or use of the Work in a manner inconsistent with the terms of this Licence Agreement. The above applies to the Work as incorporated in a Collective Work, but this does not require the Collective Work apart from the Work itself to be made subject to the terms of this Licence. If You create a Collective Work, upon notice from any Licensor You must, to the extent practicable, remove from the Collective Work any reference to such Licensor or the Original Author, as requested.

b You may not exercise any of the rights granted to You in Section 3 above in any manner that is primarily intended

for or directed toward commercial advantage or private monetary compensation. The exchange of the Work for other copyrighted works by means of digital file sharing or otherwise shall not be considered to be intended for or directed toward commercial advantage or private monetary compensation, provided there is no payment of any monetary compensation in connection with the exchange of copyrighted works.

c If you distribute, publicly display, publicly perform, or publicly digitally perform the Work or any Collective Works, you must keep intact all copyright notices for the Work and give the Original Author credit reasonable to the medium or means You are utilizing by conveying the name (or pseudonym if applicable) of the Original Author if supplied; the title of the Work if supplied. Such credit may be implemented in any reasonable manner; provided, however, that in the case of a Collective Work, at a minimum such credit will appear where any other comparable authorship credit appears and in a manner at least as prominent as such other comparable authorship credit.

5 Representations, Warranties and Disclaimer

a By offering the Work for public release under this Licence, Licensor represents and warrants that, to the best of Licensor's knowledge after reasonable inquiry:

i Licensor has secured all rights in the Work necessary to grant the licence rights hereunder and to permit the lawful exercise of the rights granted hereunder without You having any obligation to pay any royalties, compulsory licence fees, residuals or any other payments;

ii The Work does not infringe the copyright, trademark, publicity rights, common law rights or any other right of any third party or constitute defamation, invasion of privacy or other tortious injury to any third party.

b Except as expressly stated in this licence or otherwise agreed in writing or required by applicable law, the work is licenced on an 'as is' basis, without warranties of any kind, either express or implied including, without limitation, any warranties regarding the contents or accuracy of the work.

6 Limitation on Liability

Except to the extent required by applicable law, and except for damages arising from liability to a third party resulting from breach of the warranties in section 5, in no event will licensor be liable to you on any legal theory for any special, incidental, consequential, punitive or exemplary damages arising out of this licence or the use of the work, even if licensor has been advised of the possibility of such damages.

7 Termination

a This Licence and the rights granted hereunder will terminate automatically upon any breach by You of the terms of this Licence. Individuals or entities who have received Collective Works from You under this Licence, however, will not have their licences terminated provided such individuals or entities remain in full compliance with those licences. Sections 1, 2, 5, 6, 7, and 8 will survive any termination of this Licence.

b Subject to the above terms and conditions, the licence granted here is perpetual (for the duration of the applicable copyright in the Work). Notwithstanding the above, Licensor reserves the right to release the Work under different licence terms or to stop distributing the Work at any time; provided, however that any such election will not serve to withdraw this Licence (or any other licence that has been, or is required to be, granted under the terms of this Licence), and this Licence will continue in full force and effect unless terminated as stated above.

8 Miscellaneous

a Each time You distribute or publicly digitally perform the Work or a Collective Work, Demos offers to the recipient a licence to the Work on the same terms and conditions as the licence granted to You under this Licence.

b If any provision of this Licence is invalid or unenforceable under applicable law, it shall not affect the validity or enforceability of the remainder of the terms of this Licence, and without further action by the parties to this agreement, such provision shall be reformed to the minimum extent necessary to make such provision valid and enforceable.

c No term or provision of this Licence shall be deemed waived and no breach consented to unless such waiver or consent shall be in writing and signed by the party to be charged with such waiver or consent.

d This Licence constitutes the entire agreement between the parties with respect to the Work licenced here. There are no understandings, agreements or representations with respect to the Work not specified here. Licensor shall not be bound by any additional provisions that may appear in any communication from You. This Licence may not be modified without the mutual written agreement of Demos and You.

DEMOS

Demos is a champion of people, ideas and democracy. We bring people together. We bridge divides. We listen and we understand. We are practical about the problems we face, but endlessly optimistic and ambitious about our capacity, together, to overcome them.

At a crossroads in Britain's history, we need ideas for renewal, reconnection and the restoration of hope. Challenges from populism to climate change remain unsolved, and a technological revolution dawns, but the centre of politics has been intellectually paralysed. Demos will change that. We can counter the impossible promises of the political extremes, and challenge despair – by bringing to life an aspirational narrative about the future of Britain that is rooted in the hopes and ambitions of people from across our country.

Demos is an independent, educational charity, registered in England and Wales. (Charity Registration no. 1042046)

Find out more at www.demos.co.uk

DEMOS

PUBLISHED BY DEMOS DECEMBER 2023

© DEMOS. SOME RIGHTS RESERVED.

15 WHITEHALL, LONDON, SW1A 2DD

T: 020 3878 3955

HELLO@DEMOS.CO.UK

WWW.DEMOS.CO.UK