



Computational
Propaganda
Research Project

Industry Responses to Computational Propaganda and Social Media Manipulation

Emily Taylor, *Oxford Information Labs*

Stacie Hoffmann, *Oxford Information Labs*



TABLE OF CONTENTS

0	Executive Summary	3
1	Introduction	5
2	Methodology	6
3	An Overview of Major Platforms’ Responses to Computational Propaganda	8
5	Platforms’ Terms of Service, User Agreements, and Privacy and Advertising Policies	15
6	Platform Profiles	16
6.1	Facebook	16
	User Content.....	16
	Advertising Content.....	17
6.2	Google	18
	User Content.....	19
	Advertising Content.....	20
6.3	YouTube	20
	User Content.....	21
	Advertising Content.....	21
6.4	Twitter	22
	User Content.....	23
	Advertising Content.....	24
6.5	LinkedIn	25
	User Content.....	26
	Advertising Content.....	27
6.6	Reddit.....	27
	User Content.....	28
	Advertising Content.....	29
7	Conclusion	32
8	Appendices	33
8.1	Appendix 1: Examples of Country-Specific Initiatives.....	33
8.2	Appendix 2: Terms of Service Analysed	35
9	References	37
10	Series Acknowledgement	48

LIST OF TABLES

Table 1. An overview of platforms' initiatives to combat computational propaganda..	9
Table 2. Facebook – A chronology of self-regulatory responses	10
Table 3. Google – A chronology of self-regulatory responses	12
Table 4. YouTube – A chronology of self-regulatory responses	13
Table 5. Twitter – A chronology of self-regulatory responses	13
Table 6. LinkedIn – A chronology of self-regulatory responses	14
Table 7. Reddit – A chronology of self-regulatory responses	14
Table 8. Summary of platforms' user content terms and policies	30
Table 9. Summary of platforms' advertising terms and policies.....	31
Table 10. Examples of country-specific initiatives (other than USA).....	33
Table 11. LinkedIn – Terms of service	35
Table 12. Reddit – Terms of service	35
Table 13. Twitter – Terms of service.....	36
Table 14. Google – Terms of service	36
Table 15. YouTube – Terms of service	36
Table 16. Facebook – Terms of service.....	36

After the election, I made a comment that I thought the idea misinformation on Facebook changed the outcome of the election was a crazy idea. Calling that crazy was dismissive and I regret it. This is too important an issue to be dismissive.

Mark Zuckerberg in a Facebook post
27 September 2017

0 EXECUTIVE SUMMARY

The reverberations from the 2016 US presidential elections continue to be felt across the world. In November 2016, the idea that one of the world's most sophisticated democracies could have been gamed by hostile states using social media platforms seemed 'crazy' to more people than Mark Zuckerberg. But as more evidence emerges – whether from FBI investigations, announcements by the intelligence services, disclosures by whistle-blowers, the Cambridge Analytica scandal, or examples of digitally distributed disinformation in other countries and around other elections – an uncomfortable picture takes shape. Silicon Valley's leading technology companies and social media platforms have found themselves at the centre of a perfect storm.

What have Internet companies done to combat the creation and spread of computational propaganda on their platforms and services? What do the leading players' initiatives tell us about their coping strategies? How are their actions supported by the companies' terms and policies for users and advertisers? And have there been any substantial policy changes as a result of the proliferation of computational propaganda? We examined platform initiatives and terms of service agreements of six Internet companies (Facebook, Google and YouTube, LinkedIn, Reddit, and Twitter) and found:

- Immediately following the events of 2016, platforms suggested that only a low percentage of overall posts or users were involved, and therefore not many self-governing actions were taken. But by the spring of 2017, attitudes seemed to have changed and a flurry of initiatives were launched. To various degrees, different platform companies have announced the following:
 - changes to the algorithms underlying newsfeeds or ad targeting
 - new partnerships with third-party fact-checkers
 - investment in and support for quality journalism (and the business of news organizations)
 - greater transparency about electoral advertising and internal content moderation practices
 - additional investments in both automated and human content moderation.
- The initiatives that have been taken suggest some differences between the strategies of some of the largest platform companies (Facebook, Google and

YouTube, and Twitter) as they search for effective, appropriate, and credible self-regulatory responses amid a firestorm of public and political opprobrium.

- The platforms' responses also seem to be heavily influenced by news events, such as the Cambridge Analytica scandal (Facebook), the reports of Holocaust denial sites featuring prominently in search results and influencing autocomplete (Google), and research into the impact of fake accounts and bots (Twitter). Official announcements often reference current events and reporting, and their impact on companies' actions suggests that their coping strategies are still emergent at best and reactive at worst. Large technology companies used to driving change in other areas often seem to be reactive and on the back foot when it comes to combating computational propaganda.
- Overall, no major changes to terms and policies directly related to computational propaganda were observed, leading to the conclusion that current terms and policies provide plenty of opportunities to address these issues. The language of the terms and policies relating to users and advertisers tends to be widely drawn, offering flexibility for creative interpretation and different degrees and forms of enforcement. The major change indicated by the official blogs of the companies is that they have ramped up their enforcement activities, often through a combination of new automated efforts and increased investment in human content moderation.
- Finally, it is apparent that past, impending, and possibly additional regulation is having an impact on company policies and practices. New European Union (EU) steps like the General Data Protection Regulation, as well as numerous proposals for national legislation (covered by Bradshaw & Neudert (2018)) are expected to result in a raft of updates to terms and policies as well as to platforms' activities around enforcement, content moderation, etc.

The fall-out from the 2016 elections has led to public scrutiny of the business models of many of the platforms in this study and has exposed the unpredictable consequences of automated content curation, highly targeted digital advertising, and the risk that extraordinarily rich data sets fall into the wrong hands. The reactive, incremental, and piecemeal nature of the initiatives announced to date suggests that the major platform companies themselves are still not sure about how to respond effectively, appropriately, and credibly to various revelations about the impact of computational propaganda and social media manipulation on democratic processes. In the end, there may be benefits to society in keeping some types of content up or taking them down, but the key questions are who decides what to do and which process should be used, and how the workings of our media environment can be rendered more intelligible, transparent, and supportive of individual liberties and democratic politics. Ultimately, we suggest that platform companies need to work with other stakeholders to define clear, open, and transparent responses that can effectively, appropriately, and credibly help combat computational propaganda.

1 INTRODUCTION

This paper provides an inventory of the self-regulatory initiatives taken by six Internet platforms in response to computational propaganda between 1 November 2016 and 31 April 2018. It develops a chronology of self-governing initiatives and then reviews the platforms' powers to intervene arising from their standard terms and policies. The paper begins with a brief overview of the methodology used for this report. After the section that highlights the key industry responses, there is a separate section for each of the platforms which reviews their self-regulatory responses, including any variations by country, and the scope for enforcement offered by their terms of service and policies, for both users and advertisers). The paper ends with a brief conclusion.

2 METHODOLOGY

The emergence of the term ‘fake news’ in everyday language can be linked back to Donald Trump’s use of the term on the presidential campaign trail. Definitions of it are now being included in English language dictionaries. For example, the Cambridge English Dictionary defines fake news as ‘false stories that appear to be news spread on the Internet using other media, usually created to influence political views or as a joke’ (Cambridge English Dictionary, 2018). However, the term fake news lacks definitional clarity and is often used to describe a wide range of harmful content online, from purposefully distributed false stories, click bait, or propaganda to accidental misinformation or the voicing of controversial opinions. Despite the lack of clarity, government regulators have increasingly pressured social media platforms to address fake news.

Due to the lack of nuance and precision regarding the term fake news, we use the term computational propaganda to define the scope of issues that platforms are currently trying to address. Computational propaganda is the use of automation, algorithms, and big data analytics to manipulate public life (Woolley & Howard, 2016 p.4886). It is a multidimensional challenge that spans issues to do with algorithmic affordances and the spread of junk content, foreign interference operations, advertising transparency and data-driven campaigning, political bots, and the digital challenges of sustainable journalism models (Bradshaw & Neudert, 2018 p. 5). We conducted a content analysis of the official announcements of self-regulatory responses of six Internet platforms (Facebook, Google and YouTube, LinkedIn, Reddit, and Twitter) relating to computational propaganda, supplemented by selected news reports and the platforms’ relevant terms and policies.

The purpose was to create a chronology of the initiatives, discover any changes to the companies’ terms and policies since the US presidential election in 2016, and analyse the current state of those policies with reference to computational propaganda. Initiatives were reviewed between 1 November 2016 and 31 April 2018, and the terms and policies from 1 November 2016 to 31 March 2018. Due to the rapidly evolving landscape around issues involving Internet platforms, this study provides a snapshot of company initiatives, terms, and policies.

Internet companies have a variety of terms and policies, ranging from high-level user-oriented community standards to detailed legal terms. We limited our analysis to the following types of documents for each company:

- official announcements and company blogs (using keywords that include ‘fake news’, ‘misinformation’, and ‘election’)
- terms of service, including user agreements, community guidelines, privacy policies and advertising policies
- selected news reports relating to companies’ self-regulatory responses.

In total, 73 official company announcements and related news articles were examined to create the chronology of initiatives and the inventory of responses by country. For the analysis of terms, 37 policies and their relevant versions since 2016 were reviewed. Additional policies supporting or acting as supplements to those included in this study were not considered. Because the scope focused on company announcements and terms and policies, other related documents such as oral or written evidence presented to official inquiries, the FBI indictment of the Internet Research Agency, and the majority of news reports were not included but provide a rich source of contemporaneous information for future study.

Comparing terms and policies across jurisdictions was also not a focal point of this study, and only some basic information was included in this paper where relevant. As a result, only terms relevant to Europe and the UK were included here. These could be universal terms, but many companies have additional or different terms for those living in or outside the USA. This, again, is another potential area of future study.

3 AN OVERVIEW OF MAJOR PLATFORMS' RESPONSES TO COMPUTATIONAL PROPAGANDA

The 73 documents examined for this study use a variety of terms to describe the companies' responses, which often differ between different companies. To enable a comparative analysis, we have identified 10 broad categories of interventions, which are applied through the chronology and country analysis.

1. Algorithmic changes / AI – Several companies have adjusted their algorithms to demote the prominence of poor-quality news stories, or to display a range of views (to combat 'echo chamber' effects), and use AI to surface abusive behaviour, spam, or other kinds of harmful, illegal, or unwanted content.
2. Fact-checking and flagging – A number of announcements highlight partnerships with NGOs or news organizations to fact-check stories in real time. The platforms themselves have also been experimenting with flagging news – either as a result of third-party fact-checking exercises or as a result of encoded information provided by news organizations in collaboration with the platforms.
3. Enhanced reporting – Several Internet companies have begun experimenting with ways to encourage users to report poor-quality content or misinformation on their platforms.
4. Human content moderation – A number of announcements and initiatives have focused on hiring or training more staff for content moderation.
5. Enforcement – The enforcement of existing terms of service guidelines is by far the most popular response for most platforms, highlighting the central importance of companies' terms and policies and investing in enforcing them.
6. Partnerships and research – Several platforms have partnered with each other as well as with news organizations, universities, or civil society organizations to conduct research and combat computational propaganda.
7. Media literacy programmes – A long-term approach some companies have adopted is focusing on children of primary school age and other groups to help fund more teaching and training opportunities to instill the skills necessary to be able to identify fake content from fact.
8. Supporting quality journalism – To address challenges to sustainable journalism models in the digital age, some platforms have begun providing financial support for training journalists and creating tools for news organizations to help them create frictionless sign-ups to subscriptions. They have also designed interventions aimed at encouraging more diversity in newsrooms.

9. Improved transparency – To improve transparency on their platforms, some companies have announced their intention to create searchable archives of political advertisements and publicize internal policies and content moderation guidelines. Other transparency initiatives include notifying users if they have interacted with foreign operatives.
10. Data protection measures – In preparation for the GDPR and in response to growing concerns around social media and the surveillance economy, several companies strengthened their internal data and privacy protections by limiting access to data by third-party apps, APIs, and even some of their own services. The majority of initiatives under this category were announced from March 2018 onwards.

The following section provides a chronology of industry responses and how these initiatives fit into these 10 categories. There is also a country-specific collection of responses in Appendix 1.

Table 1. An overview of platforms’ initiatives to combat computational propaganda
Summary of announced initiatives by platforms, 1 November 2016–31 April 2018

Platform	Algorithm changes / AI	Fact-checking and flagging	Enhanced reporting	Human content moderation	Enforcement	Partnerships and research	Media literacy programmes	Supporting quality journalism	Improved transparency	Data protection measures
Facebook, n=39	13	6	10	9	21	12	3	5	17	6
Google, n=15	10	7	1	2	1	12	6	10	3	0
YouTube, n=2	2	1	0	1	1	1	0	0	1	0
Twitter, n=9	5	0	1	3	8	5	2	1	4	1
LinkedIn, n=4	1	0	0	1	2	0	0	0	0	0
Reddit, n=4	0	0	0	1	4	1	0	0	0	0
Total, n=73	31	14	12	17	37	31	11	16	25	7

Source: Authors, 2018. Table shows the types of initiatives announced by each of the platforms between 1 November 2016 and 31 April 2018. The numbers represent the number of times each type of initiative was mentioned (1 per announcement per category). Note that several initiatives are typically mentioned in each announcement, hence the numbers in the columns do not add up to the ‘n’ for each company. The reason for creating this numerical comparison is to give an indication of the intensity of certain types of intervention and to highlight the difference in emphasis between the companies.

Table 2. Facebook – A chronology of self-regulatory responses

Date	Title	Algorithm changes / AI	Fact-checking and flagging	Enhanced reporting	Human content moderation	Enforcement	Partnerships and research	Media literacy programmes	Supporting quality journalism	Improved transparency	Data protection measures
Nov. 2016	Zuckerberg post – intended responses to fake news and misinformation	X	X	X		X	X		X		
Dec. 2016	Newsroom – responses to fake news and misinformation	X	X			X	X				
Jan. 2017	Announcing the Facebook Journalism Project			X				X	X		
Apr. 2017	Working to stop misinformation and false news	X	X	X		X	X		X	X	
Apr. 2017	A new educational tool against misinformation						X	X			
Apr. 2017	Improvements in protecting the integrity of activity on Facebook	X				X					
Apr. 2017	Facebook and information operations									X	
Apr. 2017	Breaking new ground in the fight against fake likes	X								X	
Apr. 2017	Updates to clicks and impression reporting					X					
Aug. 2017	Taking action against video clickbait			X							
Aug. 2017	Blocking ads from pages that repeatedly share false news					X					
Sept. 2017	An update on information operations on Facebook					X					
Sept. 2017	Joining the Media Ratings Council and Trustworthy Accountability Group; adding 3,000 content reviewers				X		X				
Sept. 2017	Facebook to provide Congress with ads linked to Internet Research Agency									X	
Sept. 2017	Updating our Ad Targeting					X					
Sept. 2017	Zuckerberg post					X	X			X	
Oct. 2017	New test to provide context about ads			X			X		X		
Oct. 2017	Improving enforcement and transparency of ads (Russia)	X			X	X				X	
Oct. 2017	Ads transparency and authenticity efforts			X						X	
Nov. 2017	Continuing transparency on Russian activity			X							

Date	Title	Algorithm changes / AI	Fact-checking and flagging	Enhanced reporting	Human content moderation	Enforcement	Partnerships and research	Media literacy programmes	Supporting quality journalism	Improved transparency	Data protection measures
Nov. 2017	Our advertising principles	X				X				X	
Dec. 2017	Fighting engagement bait on Facebook	X									
Dec. 2017	Restrictions on detailed targeting of groups based on sensitive criteria					X					
Dec. 2017	New tool tells users if they have liked or followed Russia's troll army									X	
Dec. 2017	Replacing disputed flags with related articles	X	X		X			X			
Jan. 2018	Hard questions: what effect does social media have on democracy		X	X	X	X	X			X	
Jan. 2018	Helping ensure news on Facebook is from trusted sources	X									
Mar. 2018	Hard questions: what is Facebook doing to protect election security?		X		X	X	X		X	X	
Mar. 2018	Suspending Cambridge Analytica and SCL from Facebook				X	X					X
Mar. 2018	Pursuing forensic audits to investigate Cambridge Analytica claims					X					X
Mar. 2018	Cracking down on platform abuse					X				X	X
Apr. 2018	Authenticity matters: the IRA has no place on Facebook	X			X	X					
Apr. 2018	An update on our plans to restrict data access on Facebook					X				X	X
Apr. 2018	Hard questions: Q&A with Mark Zuckerberg on protecting people's information	X			X	X					X
Apr. 2018	Making ads and pages more transparent			X		X	X			X	
Apr. 2018	Facebook launches new initiative to help scholars assess social media's impact on elections						X				
Apr. 2018	Hard questions: what information do Facebook advertisers know about me?									X	X
Apr. 2018	Publishing our internal enforcement guidelines and expanding our appeals process	X		X	X		X			X	
Apr. 2018	Hard questions: why do you leave up some posts but take down others?									X	

Source: Authors, 2018. Data collected from Zuckerberg (2016), Mosseri (2016), Simo (2017), Mosseri (2017b), Mosseri (2017a), Facebook Security (2017), Weedon, Nuland, & Stamos (2017), Cevahir (2017), Facebook Newsroom (2017a), Hamodi, Bokhari, & Zhang (2017), Shukla & Lyons (2017), Stamos (2017), Everson (2017), Stretch (2017), Facebook Business (2017), Zuckerberg (2017), Anker, Su, & Smith (2017), Kaplan (2017), Goldman (2017a), Facebook Newsroom (2017b), Goldman (2017b), Silverman & Huang (2017), Bala (2017), Lyons (2017), Chakrabarti (2018), Mosseri (2018), Facebook Newsroom (2018a), Grewal (2018), Facebook Newsroom (2018b), Facebook Newsroom (2018c), Stamos (2018), Schroepfer (2018), Facebook Newsroom (2018d), Goldman (2018a), Schrage & Ginsberg (2018), Goldman (2018b), Bickert (2018b), Bickert (2018a).

Table 3. Google – A chronology of self-regulatory responses

Date	Title	Algorithm changes /AI	Fact-checking and flagging	Enhanced content reporting	Human moderation	Enforcement	Partnerships and research	Media literacy programmes	Supporting quality journalism	Improved transparency	Data protection measures
Apr. 2017	Our latest quality improvements for search	X			X						
Apr. 2017	Internet citizens: let's make a better web						X	X			
Apr. 2017	Fact Check now available in Google Search and News Around the World		X				X				
May 2017	Launch of 'Be Internet Awesome'						X	X			
Jun. 2017	Developing critical reading skills with media literacy apps on Chromebooks							X			
Jul. 2017	Making it easier for publishers to share fact-checked content		X						X		
Oct. 2017	Security and disinformation in the US election						X			X	
Nov. 2017	Fact-checking the French election: lessons from CrossCheck, a collaborative effort to combat misinformation		X				X		X		
Nov. 2017	Identifying credible content online, with help from the Trust Project	X					X		X		
Dec. 2017	News Lab in 2017: the year in review	X	X	X			X		X		
Dec. 2017	News Lab in 2017: working with news organizations to address industry challenges	X	X				X	X	X		
Mar. 2018	The Google News Initiative: building a stronger future for news	X					X	X	X		
Mar. 2018	Introducing Subscribe with Google								X		
Mar. 2018	Elevating quality journalism on the open web	X	X				X	X	X	X	

Source: Authors, 2018. Data collected from Gomes (2017), Grummer (2017), Kosslyn & Yu (2017), Google (2018d), Greenleaf (2017), Anderson (2017), Walker & Salgado (2017), Gingras (2018), DoubleClick (2017), Dieudonne (2017), Chang (2017), Grove (2017b), Grove (2017a), Schindler (2018), Albrecht (2018), Gingras (2018).

Table 4. YouTube – A chronology of self-regulatory responses

Date	Title	Algorithm changes / AI	Fact-checking and flagging	Enhanced reporting	Human content moderation	Enforcement	Partnerships and research	Media literacy programmes	Supporting quality journalism	Improved transparency	Data protection measures
Dec. 2017	YouTube: expanding our work against abuse of our platform	X			X	X	X			X	
Feb. 2018	YouTube: greater transparency for users around news broadcasters	X	X								

Source: Authors, 2018. Data collected from YouTube (2017b), YouTube (2018d).

Table 5. Twitter – A chronology of self-regulatory responses

Date	Title	Algorithm changes / AI	Fact-checking and flagging	Enhanced reporting	Human content moderation	Enforcement	Partnerships and research	Media literacy programmes	Supporting quality journalism	Improved transparency	Data protection measures
Jun. 2017	Our approach to bots and misinformation	X			X	X	X				
Sept. 2017	Update: Russian interference in the 2016 US election, bots, and misinformation	X				X		X		X	
Oct. 2017	Announcement: RT and Sputnik advertising					X	X				
Oct. 2017	U.S. Senate Committee on the Judiciary: opening remarks	X			X	X				X	
Oct. 2017	New transparency for ads on Twitter			X		X	X			X	
Jan. 2018	Update on Twitter's review of the 2016 US election	X				X	X	X	X	X	
Feb. 2018	Automation and the use of multiple accounts					X					X
Mar. 2018	Twitter Health Metrics Proposal Submission						X				
Apr. 2018	Serving the public conversation during breaking events	X			X	X					

Source: Authors, 2018. Data collected from Crowel (2017), Twitter Public Policy (2017a), Twitter Public Policy (2017b), Edgett (2017), Falck (2017), Twitter Public Policy (2018), Roth (2018), Harvey (2018).

Table 6. LinkedIn – A chronology of self-regulatory responses

Date	Title	Algorithm changes / AI	Fact-checking and flagging	Enhanced reporting	Human content moderation	Enforcement	Partnerships and research	Media literacy programmes	Supporting quality journalism	Improved transparency	Data protection measures
Dec. 2016	Fake LinkedIn profiles used by hackers					X					
Dec. 2016	Daniel Roth interview				X						
Mar. 2017	Introducing Trending Storylines	X									
Mar. 2018	Updates to our Terms and Services					X					X

Source: Authors, 2018. Data collected from BBC (2015), Johnson (2016), Cohen (2017), Harrington (2018)

Table 7. Reddit – A chronology of self-regulatory responses

Date	Title	Algorithm changes / AI	Fact-checking and flagging	Enhanced reporting	Human content moderation	Enforcement	Partnerships and research	Media literacy programmes	Supporting quality journalism	Improved transparency	Data protection measures
Mar. 2018	Russian propaganda remains on Reddit					X					
Mar. 2018	Reddit and the struggle to detoxify the Internet					X					
Apr. 2018	Reddit's 2017 transparency report and suspect account findings					X					
Apr. 2018	In response to recent reports about the integrity of Reddit, I'd like to share our thinking				X	X	X				

Source: Authors, 2018. Data collected from Lapowsky (2018), Marantz (2018), Reddit (2018e), Huffmann (2018).

5 PLATFORMS' TERMS OF SERVICE, USER AGREEMENTS, AND PRIVACY AND ADVERTISING POLICIES

Content moderation, terms of service, and advertising and privacy policies form the basis of how platforms structure the governance of user activities. Thus, when evaluating industry initiatives to address computational propaganda, it would make sense to determine whether any key changes were made to the core rules and guidelines of the major Internet platforms. Open discussion of the platforms' content moderation activities is limited, especially for Google, Twitter, and others. Even prior to 2016, studies and articles have revealed extensive activity by many Internet platforms in this regard (Chen, 2014; Roberts, 2016; Taylor, 2016, p.11) by a large workforce of humans. But due to the amount of content being posted every second, moderation is also automated through the use of machine-learning and artificial intelligence technologies. Some of the platforms are beginning to be more transparent about these processes. Both Google and Facebook have published their internal guidelines for content moderation over the past 18 months. Facebook has also been more open about the number of staff it employs to review content and has introduced a right of appeal for some decisions. However, there is still more work to be done to improve transparency about how these major decisions are made and how often they are made. The following section analyses how computational propaganda can be or has been addressed in the policies of social media platforms. It is organized into sections on each of the six platforms discussed in this paper. At the end of the section, Tables 8 and 9 are summary tables for (1) user and content policies and (2) advertising policies respectively. It is also worth noting that availability of previous and future terms varies by company. In instances where previous versions are not available, analysis is based on the current terms.

6 PLATFORM PROFILES

6.1 Facebook

Of all the companies in this study, Facebook has borne the brunt of public attention since the 2016 elections. Following revelations that some users and advertisers were abusing the social media platform to circulate propaganda, disinformation, and dark advertisements, governments held evidence sessions and executed formal inquiries into the platform. Facebook's Mark Zuckerberg initially dismissed as 'crazy' the idea that the Russians could have used Facebook to affect the outcome of the US elections. Since then, between November 2016 to April 2018, the company produced at least 39 official announcements of initiatives (see Table 2). Facebook's most frequently referenced intervention is enforcement of its terms and policies, against both users and advertisers. For example, it took action against over 30,000 fake accounts in the run-up to the French presidential elections in 2017 (Facebook, 2017). It also increased its human content moderation team by approximately 5,000 in 2018 (Rushe, 2018).

The chronology of Facebook's self-regulatory responses (see Table 2) reveals a shifting, emergent strategy which often seems affected by news reporting. Following the Cambridge Analytica scandal there was a marked shift towards banning individuals who break their terms of service (Grewal, 2018) and locking down third-party access through apps and APIs, while emphasizing user privacy and security (Schroepfer, 2018). There have also been a few country-specific interventions, including additional flagging tools for controversial content in Germany (Oltermann, 2018) and the fact-checking initiative CrossCheck in France (Rosen, 2018).

User Content

Facebook is the only company included in this study which did not offer any versioning of its policies. As a result, it is difficult to draw conclusions about relevant changes after the 2016 elections or preparation for the new NetzDG laws. Since concluding research for this study, Facebook has made available new versions of its Statement of Rights and Responsibilities (now called Terms of Service) and Data Policies – both soon to come into effect – and has updated its Community Standards. However, given the time of writing it was not possible to include an analysis of these versions here.

Facebook's Community Standards are referred to in multiple terms and policies as the foundational policy document (Facebook, 2018c). The Standards are broad and high level, requiring readers to follow a winding trail of supplementary documents to find more detail about the terms and policies. Of all the relevant sections included in the Community Standards (listed below), none include direct reference to computational propaganda, fake news, propaganda, disinformation, or misinformation (Facebook, 2018c):

- Direct Threats
- Self Injury
- Dangerous Organisations
- Bullying and Harassment
- Attacks on Public Figures
- Criminal Activity
- Sexual Violence and Exploitation
- Nudity
- Hate Speech
- Violence and Graphic Content
- Using Your Authentic Identity
- Fraud and Spam
- Intellectual Property

Instead, content policies across Facebook's documents tend to focus on violent, harmful, and sexually explicit content as well as some vague categories. For example, the Community Standards prohibits 'offensive, inappropriate, obscene, ... or otherwise objectionable content' that might not be illegal but could be considered undesirable (Facebook, 2018b). The Statement of Rights and Responsibilities also prohibits 'misleading, malicious or discriminatory' actions, but only with reference to safety (Facebook, 2018c).

The vague language highlights the tension that Internet companies contend with. Reluctant to become 'arbiters of truth', they need to have the right to remove content that may cause harm to its brand as a safe, trusted online space (Chakrabarti, 2018). These decisions are not easy: for example, the removal of the iconic 'napalm girl' photo in 2016 was incredibly contentious (Levin, Wong, & Harding, 2016). Facebook acknowledges the difficulty of analysing content, particularly that which may be 'humour, satire, or social commentary' (Facebook, 2018b), highlighting some of the reasons why determining socially acceptable content is not a task that should fall solely to the Internet companies.

In Facebook's Community Standards, the use of 'misleading or inaccurate information' is specifically with reference to fraud and spam and the use of such information to artificially collect 'likes' is also prohibited (Facebook, 2018b). Similar to other companies included in this study, Facebook's policy on spam and fraud enables it to take action against elements of computational propaganda.

Advertising Content

Facebook's Advertising Policies, like those of many companies in this study, are detailed but do not explicitly address computational propaganda. Although there are prohibitions on 'misleading or false', 'sensational', 'controversial', and 'shocking' ads, Facebook's advertising policy primarily focuses on typical consumer protections (i.e., highly regulated products or services such as financial services, weapons, and

drugs), sex, and violence. There are five key advertising policy areas (Facebook, 2018a):

- Prohibited content
- Restricted content
- Targeting
- Lead adverts
- Data use restrictions.

Controversial content is the only area to reference politics and to provide a mechanism to take down clickbait, as it states: ‘Adverts must not contain content that exploits controversial political or social issues for commercial purposes’ (Facebook, 2018a). According to the policies, advertisements also go through a review process to ‘check [the] advert’s images, text, targeting and positioning, in addition to the content on your advert’s landing page’ (Facebook, 2018a). Facebook claims that ‘most’ advertisements are reviewed within 24 hours; however, it is unclear whether this review is conducted by a human moderator or a machine, as this distinction is not clearly stated in the policies.

As of March 2018, Facebook’s policies make only two explicit prohibitions related to politics and advertising (Facebook, 2018a):

- Adverts must not contain content that exploits controversial political or social issues for commercial purposes.
- Adverts must not request information regarding political affiliation without our prior permission.

Finally, in Mark Zuckerberg’s question session with the US Congress in April 2018, he stated that Facebook had been working with the Senate on the Honest Ads Act, which will put in place new policy regarding online political advertising (Washington Post, 2018). If adopted by the US government, the new legislation is likely to have a direct impact on Facebook’s terms related to political advertising.

6.2 Google

Google’s response to fake news contrasts markedly with that of Facebook. While Facebook has activity across all 10 of the categories in this study, Google’s strategy appears to position the company as a neutral conduit that focuses its efforts on creating technical tools and using AI to tweak its algorithms while partnering with and funding appropriate experts and programmes to improve media literacy and strengthen quality journalism (see Table 3).

Across 16 company announcements, there are few references to human content moderation (despite acknowledging a workforce of 10,000 moderators for YouTube), enforcement of terms, or data protection measures (YouTube, 2017). Instead, Google often creates new products to solve specific problems – such as tools to help

newspapers gain subscribers (Albrecht, 2018) – or offers services such as advanced media literacy training provided to 20,000 students through an online tool (Gingras, 2018).

Like Facebook, Google displays a sensitivity to news reporting; it announced adjustments to its autocomplete and ranking algorithms following revelations that Holocaust denial sites were among the top search results (Pasquale, 2017). Country-specific initiatives include media literacy programmes in the UK, Brazil, and Canada (Gingras, 2018) and funding for teaching Canadian students how to spot fake news (Toronto Sun, 2017). And along with Facebook, Google has sponsored Media Smarts, Canada's Centre for Digital and Media Literacy (MediaSmarts, 2018).

Google's policies are meant to form the baseline for its services, with additional terms and policies specific to those services. For the purpose of this study, Google's general terms and policies were analysed. Terms and policies for individual services provided by Google, such as Google Drive, Gmail, Google+, and Google Search (with the exception of YouTube), were considered out of the scope for this study but warrant a focused analysis.

Of all the companies included in this study, Google has the most detailed advertising and ad content policies. They include sections on 'misrepresentation', 'misleading content', and 'political advertising', as well as personalized advertising with political content (Google, 2018c). Google also has more rights than other companies included in this study to remove user content. In addition to content that is 'illegal, content that ... violates [Google's] policies' may be removed. Yet the terms lack clear definitions as to what constitutes unacceptable content beyond what is clearly illegal.

Google's Terms of Service, Privacy Policy, and AdWords Advertising Policy were all updated between November 2016 and March 2018. However, there is little to link those changes with so-called fake news or foreign meddling in national affairs. Instead, like the alterations made by other companies in this study, most changes were made to reflect compliance with existing policies and regulations, particularly the GDPR and the EU-US/Swiss-US Privacy Shield frameworks. Unlike the other platforms, which usually name specific jurisdictions such as the EU, Google's policies reference abiding by national laws regarding disputes, consumer rights, and applicability of national law where California law is not applicable (Google, 2018e).

User Content

Google's policies employ the broadest language relevant to user content of all the companies included in this study. There are no references to fake news, propaganda, dis- or misinformation in the documents; references are only made to 'illegal' content or that which violates its policies (Google, 2018e). Google also states that 'automated systems analyse your content (including emails)' to provide tailored content and services, but does not mention using it to identify or address fake accounts or news, or illegal content (Google, 2018e). In addition, the company

distances itself from content liability, stating in the Terms of Service that ‘we don’t make any comments about the content within the services’ (Google, 2018e).

Advertising Content

The Advertising – Privacy & Terms policy provides that violators of Google’s policy will have their accounts terminated (Google, 2018b). While the short list of reasons for termination does not expressly relate to computational propaganda, the AdWords policy provides sufficient legal cover. Google’s AdWords Advertising Policy is the longest advertising policy reviewed for this study and includes a list of country-specific requirements for advertising – the only resource of its type found during this study. The sections and topics in the policy that are pertinent to addressing various aspects of computational propaganda include the following (Google, 2018c):

- Misrepresentation
- Misleading content
- Political content and affiliation
- Inappropriate content
- Personalized advertising.

The section on ‘misleading content’ is most relevant in the context of computational propaganda (Google, 2018c). Misleading content includes false statements about identity, and Google is the only company to mention misleading affiliations with other individuals or organizations such as government agencies or coopting related seals (Google, 2018c). The latter may apply in enforcement against computational propaganda if, for example, an advertiser is claiming or implying affiliation with a political party, candidate, or political action committee.

The AdWords policy also includes a number of specifics related to political advertising, including soliciting donations, political affiliation and content in personalized advertising, and political content generally (Google, 2018c). All of the political advertising policies are subject to local laws. For example, Google does not allow political affiliation to be used in personalized advertising or Gmail advertising, with the exception of the United States, where this is allowed. Google AdWords also does not allow what it calls ‘hyper targeting of users’ to avoid ‘an overly narrow or specific audience’ (Google, 2018c). However, this only applies to retargeting advertising (ads directed towards users that have engaged with the advertiser previously), not first-round advertisements.

6.3 YouTube

The chronology of self-regulatory responses for YouTube includes only three identified initiatives: one provides detail on the work of human content moderators (YouTube, 2017a), another announces the roll-out of ‘notices’ below publicly funded or government-funded videos (YouTube, 2018c), and the third introduces a ‘Breaking News’ shelf on the homepage in YouTube search results that will highlight news from

authoritative sources (YouTube, 2018c). These changes came about amid concerns that auto-play can potentially introduce users to ever-more-extreme content (Tufekci, 2018). No country-specific initiatives relating to YouTube had been found at the time of writing.

User Content

Unlike Google – YouTube’s parent company – the previous versions of YouTube’s terms and policies are not published. The Community Guidelines underlying YouTube’s policies provide more information about expectations of content. They are also more extensive than Google’s terms, which focus mainly on ‘illegal’ content. Instead, YouTube’s focus is on ‘shocking’ content (such as violent, graphic, or harmful content), terrorism, hate, sexual content, spam, and infringement of rights such as copyright and privacy (YouTube, 2018b).

In the YouTube policies, more detailed content violations are associated with specific content categories, which limits the flexibility of these terms and the ability to apply them to certain kinds of computational propaganda. For example, ‘illegal content’ is primarily associated with harmful or dangerous content (YouTube, 2018b). Reference to ‘deceptive’ or ‘misleading’ content is associated with spam and scams, while ‘fake’ content is associated only with violent content (YouTube, 2018b). Similar to the stance of other companies studied here, the association of computational propaganda with spam provides a less controversial and more attractive avenue for enforcement. Here, the definition of spam includes driving traffic away from YouTube (such as clickbait) and artificially ‘[increasing] the number of views, likes, comments or any other metric’ (YouTube, 2018b). YouTube stresses the idea of context in relation to content more than the other companies included in this study. Context is seen as a tool to help human moderators decide on appropriate content and potential action – such as content removal or blocking.

Finally, YouTube’s policies focus on the ‘flagging’ aspect of their content review and moderation process. YouTube states that it employs both human and machine-learning techniques in an ‘effort to improve the enforcement of [its] policies’ (YouTube, 2018b). However, it is unclear how these efforts are put into practice or how the human and machine moderators influence or interact with one another.

Advertising Content

Although AdSense¹ is a ‘Google policy’, it is considered here because YouTube lists it along with Advertiser Friendly Content Guidelines as relevant policy for content creators and advertisers. Similar to YouTube’s and Google’s other policies, the AdSense content policy focuses on ‘shocking’ or ‘illegal’ content (Google, 2018a). It

¹ AdSense differs from Google’s AdWords in that AdSense oversees the delivery of advertisements to ‘individuals’ websites’, while AdWords governs the delivery of advertisements to Google Search and other partner sites.

also discusses further the practice of artificially ‘inflating impressions and/or clicks’ (Google, 2018a).

The Advertiser Friendly Content Guidelines, on the other hand, offer more insight into attempts at content moderation (YouTube, 2018a). There is strong discouragement of videos containing ‘potentially controversial or offensive content’ (YouTube, 2018a). Topics that might fall into this category include:

- Controversial issues and sensitive events
- Harmful or dangerous acts
- Hateful content
- Inappropriate language
- Inappropriate use of family entertainment characters
- Incendiary and demeaning content
- Sexually suggestive content
- Violence.

However, these topics do not cover all of the elements of computational propaganda and junk news, such as misinformation, fake news, and disinformation.

6.4 Twitter

Twitter made a total of nine announcements with respect to issues related to computational propaganda. Overall, Twitter’s emphasis has been on enforcement of its terms and policies, primarily invoking anti-spam mechanisms to combat misinformation on its platform. Like Facebook and Google, Twitter’s self-regulatory responses also highlight adjustments to its algorithms to limit the visibility of low-quality content, and the use of AI to identify bots. Twitter also displays a sensitivity to news and research, apparently deleting millions of followers from celebrity accounts days after revelations about a proliferation of fake accounts (Confessore, Dance, & Harris, 2018). No country-specific approaches from Twitter were identified.

Since 2016, no substantial changes to Twitter’s policies can be easily linked to computational propaganda. It is expected that Twitter’s Ads Policies will be updated in the near future as the company has reported that it will be making policy changes regarding advertising and transparency, particularly political advertising (Falck, 2017). Twitter is also doing work outside its terms and policies to meet contemporary challenges. For example, days after the shooting outside YouTube headquarters on 3 April 2018, Twitter released a blog post about how it had ‘refined [its] tools, improved the speed of [its] response, and identified areas where [it] can improve’ when responding ‘to people who are deliberately manipulating the conversation on Twitter’ (Harvey, 2018). Notably, this was in response to what Twitter calls in the blog post ‘deceptive, malicious information’, as opposed to ‘credible and relevant information’. Twitter also stated that it was able to apply existing policy areas to address the issue, including:

- rules on abusive behavior
- hateful conduct policy
- violent threats
- rules against spam.

When implementing these policies, Twitter was able to take actions including:

- ‘Requiring account owners to remove Tweets’
- ‘Suspended hundreds of accounts’
- ‘Implement proactive, automated systems to prevent people who had been previously suspended from creating additional accounts to spam or harass others’
- Use the proactive, automated systems to ‘surface potentially violating Tweets and accounts to our team for review’.

User Content

Twitter has two full sets of terms of service – one for those living in the USA and another for those living outside it (Twitter, 2018d). The terms for those living *outside* the USA were analysed for the purpose of this study, and updates to those terms came into effect on 2 October 2017. The first sentence of The Twitter Rules sets out its belief that people should have ‘the power to create and share ideas and information instantly, without barriers’ (Twitter, 2018d). Further on, the rules directly reference ‘freedom of expression’ in support of open dialogue – one of the only companies to directly reference human rights in its terms. The Twitter Rules are organized into four distinct categories:

- Content Boundaries and Use of Twitter
- Abusive Behavior
- Spam and Security
- Content Visibility

Three of these four categories have information relevant to computational propaganda. ‘Spam and Security’ lists fake accounts as a tool for identifying spam-related behaviour. And impersonation, considered under ‘Abusive Behavior’, is not allowed if the intent is to mislead, confuse, or deceive. The section on ‘Content Boundaries and Use of Twitter’ is separated into the following six categories:

- Intellectual property
- Graphic violence and adult content
- Unlawful use
- Trends
- Misuse of Twitter badges
- Misuse of usernames.

There are no explicit references to elements of computational propaganda in the rules on content boundaries. The majority of changes recently made to Twitter's policies on content boundaries focus on the direct removal of content. For example, the definition 'content that violates the User Agreement, including for example, copyright or trademark violations, impersonation, unlawful conduct, or harassment' was added to Twitter's Terms of Service in October 2017 (Twitter, 2017). Twitter's Privacy Policy also went through a round of changes in June 2017 (Twitter, 2018c). While this was a full year ahead of the GDPR coming into effect, the changes reflect a move towards GDPR-compliant policy, such as user data controls and details of what data about them is collected and kept.

Twitter acknowledges in its Terms of Service that the company does not 'endorse, support, represent or guarantee the completeness, truthfulness, accuracy, or reliability of any Content or communications' (Twitter, 2017). Furthermore, the user enters into an agreement whereby they 'understand that by using the Services, [they] may be exposed to Content that might be offensive, harmful, inaccurate or otherwise inappropriate, or in some cases, postings that have been mislabeled or are otherwise deceptive' (Twitter, 2017). Although strong statements like this which act to separate the company from content and liability – and are often relevant to fake news – are regularly found in Internet companies' terms and policies, language referencing 'truthful' content in any form is something that most companies in this study have not adopted. In fact, only LinkedIn (LinkedIn, 2018d), Reddit (Reddit, 2018a), and Twitter used language related to truth (i.e., truth, truthful, untruthful) in any of their policies.

Advertising Content

Twitter Ads Policies are high level but prohibit and restrict certain kinds of advertisements. The policies employ the terms 'authenticity' and 'honest' to encourage appropriate content (Twitter, 2018e). The two categories of prohibited content are 'hateful content' and 'inappropriate content'. Inappropriate content is considered content that is

- Threatening, violent, gruesome, abusive, shocking, disturbing, offensive, vulgar, and obscene
- 'Inflammatory or provocative content which is likely to evoke a strong negative reaction
- Inappropriate targeting upon or after a sensitive event, including but not limited to: natural disasters, high profile deaths, or political uprisings
- The glorification of self-harm or related content' (Twitter, 2018e).

As with other companies, there is a strong focus on restraining violent, harmful, or sexual content in the Twitter Ads Policies. But Twitter's advertising policies also include broad terms that could be implemented with reference to computational propaganda such as 'shocking', 'offensive', 'inflammatory or provocative' (Twitter,

2018a). The policies also include a resource for best practices for advertisers, but this document was not reviewed as part of this study (Twitter, 2018b).

Political campaigning is listed as a ‘restricted’ promotional product or service, but the terms do not describe what types of or how ‘restrictions’ are applied, apart from requiring that ‘political advertisers must also comply with any applicable laws’ (Twitter, 2018e). In addition, Twitter Ads Master Service Agreement does not include terms on political advertising, but does use additional terms and language to those found in the Twitter Ads Policies that may be relevant to taking down computational propaganda, particularly in article 8, which states an agreement not to ‘engage in, nor cause others to engage in, spamming or improper, malicious, or fraudulent clicking, impression, or marketing activities relating to any Program’ (Twitter, 2018f). In the chronology of Twitter initiatives (see Table 5), it is clear that this term appears to be most frequently relied upon to address the monetization of computational propaganda. In addition, advertisers must agree not to use materials that are

- false
- deceptive
- misleading
- defamatory
- libellous.

As seen here, Twitter is among the companies included in this study which has more specific requirements regarding advertising content than the user-generated content. However, the majority of the discussion around the 2016 US election is related to user accounts and content, not advertising.

6.5 LinkedIn

For the most part, LinkedIn has avoided direct criticism about computational propaganda since the 2016 US elections. This could be related to its specific policies on political advertising, or the fact that LinkedIn bills itself as a *professional* social network, which tends towards a different audience and type of content. For instance, the Professional Community Guidelines (soon to be Professional Community Policies) state that the ‘Services shouldn’t be used to harm others or their career or business prospects’ (LinkedIn, 2018b). This is further reflected in LinkedIn’s policies, which focus on users’ ‘real’ professional persona. Although we did not identify any official company materials relating to computational propaganda, there have been news reports of LinkedIn taking down fake accounts uncovered by Symantec (BBC, 2015). For comparison, LinkedIn provides two versions of its Professional Community Guidelines (Policies) and Privacy Policy. LinkedIn made a couple of notable changes to its User Agreement that came into effect in May 2018. These are the requirement to add a reference to ‘false information’ and a section on ‘automated processing’ (LinkedIn, 2018d).

Overall, it appears that LinkedIn has fairly relevant terms and policy language related to computational propaganda, such as rules regarding real people and accounts as well as inaccurate and untruthful language. However, policies such as these may not be practical for all platforms. This is particularly true for those which allow users to have an online identity that may not be a direct reflection of their offline personality (i.e., through the use of pseudonyms). Apart from LinkedIn, Facebook is the only known company in this study to have a real-names policy.

User Content

LinkedIn's updated User Agreement, Privacy Policy, and Professional Community Policy coincided with the introduction of the GDPR, and the language added includes points on 'governing law' (LinkedIn, 2018d), the identification of 'designated countries', information on data collection, and user control rights (LinkedIn, 2018c). More relevant to computational propaganda is the inclusion of the words 'false information' in the new User Agreement and Professional Community Policies.

The Professional Community Policies also include a new reference to 'false information' as a part of the 'Be Professional' guideline (LinkedIn, 2018b), which states:

When creating content on the Services, we ask that you follow our Publishing Platform Guidelines and keep your content professional, relevant, and on-topic. Do not share false information or use the Services to shock or intimidate others. And it is not okay to share graphic, obscene, or pornographic content on the Services (LinkedIn, 2018b).

Language related to 'shocking', 'graphic', and sexual content is common in terms of service agreements. But LinkedIn also uses language related to truth, trickery, and false information throughout various policies (i.e., users may not 'use the Services to commit fraud or try to trick others') (LinkedIn, 2018b). In the User Agreement, 'false information' is used in relation to creating user accounts, where 'information will be truthful' and 'misrepresenting' identity is not allowed (LinkedIn, 2018d). In simple terms, the user should be who they say they are, supporting LinkedIn's branding as a professional platform. LinkedIn is trying to avoid trolls and fake accounts being used on its platform to disseminate and perpetuate junk news and disinformation, and its reliance on policies that insist on real names and real people offers it a justification for removing user accounts or content if they are in violation of the policies.

Additional content that is in violation of LinkedIn's User Agreement includes 'being dishonest, abusive, or discriminatory' and the posting of 'inaccurate, defamatory, obscene, shocking hateful, threatening, or otherwise inappropriate content or airing personal grievances or disputes' (LinkedIn, 2018d). Furthermore, the Publishing Platform Guidelines prohibit misleading, obscene, and discriminatory language. The company also distances itself from user-generated content and liability by noting that

‘content or information ... might be inaccurate, incomplete, delayed, misleading, illegal, offensive or otherwise harmful’ (LinkedIn, 2018b).

As stated in the current User Agreement, LinkedIn ‘does not review content’ on its platforms (LinkedIn, 2018d). This leads to the assumption that LinkedIn relies on reporting mechanisms embedded in the platform to flag and take down content in violation of the terms. Content can be reported as ‘inappropriate’, ‘fake, spam, scam’, or ‘something else’. So reporting mechanisms do allow a direct method for users to report what is believed to be computational propaganda.

Advertising Content

LinkedIn’s Advertising Guidelines include requirements that content ‘should not be fraudulent, deceptive or include misleading titles, statements, or illustrations’ and should not include lies, exaggerations, or false claims (LinkedIn, 2018a). Additionally, the guidelines prohibit ‘provocative speech’, which includes ‘speech related to political or religious topics’ (LinkedIn, 2018a). Political advertisement guidelines include identifying ‘the person or entity that paid for the message’, indication of whether a candidate has authorized a message, contact information (in some cases), and must ‘comply with applicable laws’ such as identifying sponsors and observing blackout periods. However, guidelines on political advertisements do not touch on content or expectations that ‘inaccurate’, ‘misleading’, or ‘false information’ should not be used – although such expectations are relevant to user-generated content.

6.6 Reddit

In March 2017, leaked documents reported that the Internet Research Agency (IRA) – reportedly a Kremlin-backed organization – deployed trolls on Reddit and 9Gag with the intention of influencing the US presidential campaign (Collins & Russell, 2018). Soon after, a post by Reddit CEO and co-founder Steve Huffman confirmed that Russian propaganda had been found on the platform (Huffman, 2018). These actions resulted in the platform being included in the US Senate’s investigation into Russian meddling in the 2016 US elections (Romm, 2018).

Reddit has made a limited number of official statements in relation to its self-regulatory responses. Some initiatives include the removal of profiles related to the IRA and blocking ads by Russian entities. No country-specific initiatives were found.

As outlined in its User Agreement, Reddit may be more tolerant of content that is ‘funny, serious, offensive, or anywhere in between’ (Reddit, 2018c) to ‘encourage a fair and tolerant place for ideas, people, links, and discussion’ (Reddit, 2018f). The Content Policy outlines a space for ‘authentic content’ and discussion, but notes that there might be threads and forums that promote extreme views, violence, and lewd or offensive content – many denoted by the ‘NSFW’ (Not Safe for Work) tag.

In addition to Reddit's high-level User Agreement and Content Policy, the platform also relies on 'moderators', or users who set their own rules for their threads, ensure the thread complies with Reddit policies, and manage disputes between Redditors. The Monitor Guidelines for Healthy Communities allow both 'discussion (and appeal) of moderator actions' (Reddit, 2018d). Thus, Reddit staff will only get involved when there is poor moderation or if content breaks their high-level policies.

Most of the language conveying expectations of content in Reddit policies is in the Advertising Policy, including expectations for landing pages (Reddit, 2018a). This element of Reddit's policy may help to discourage the commodification of low-quality or junk content, such as clickbait, or propaganda.

User Content

Although there were several updates to the Reddit User Agreement and Reddit Rules between the 2016 elections and March 2018, updates were only made with reference to copyright infringement and reporting (Reddit, 2018f). Changes to Reddit's Privacy Policy since the 2016 elections have been minimal and were most likely in preparation for the GDPR (Reddit, 2017). Like many companies, updates to the Privacy Policy focus on clarifying which data is collected and how that is done, for what purpose it is collected, and the extent of user controls. New sections include 'information collected from integrations', 'location information', and 'information we collect from other sources' (Reddit, 2017).

Reddit's Content Policy focuses on content that is violent, sexual, or harmful in nature (Reddit, 2018c). Categories flagged as 'prohibited' in the policy include content that

- Is illegal
- Is involuntary pornography
- Is sexual or suggestive content involving minors
- Encourages or incites violence
- Threatens, harasses, or bullies or encourages others to do so
- Is personal and confidential information
- Impersonates someone in a misleading or deceptive manner
- Uses Reddit to solicit or facilitate any transaction or gift involving certain goods and services
- Is spam.

Illegal content, impersonations, and spam are the most obvious categories to associate with computational propaganda. However, the only section of the Content Policy which mentions 'fake' content is with reference to 'involuntary pornography' (Reddit, 2018c). And the detail regarding impersonations includes the only reference to 'misleading' or 'deceptive' content (Reddit, 2018c).

Another relevant aspect of Reddit's terms is spam, which is broadly defined by the Reddit Content Policy as 'repeated, unwanted, and/or unsolicited actions, whether automated or manual, that negatively affect Reddit users, Reddit communities, and/or Reddit itself' (Reddit, 2018c). This includes actions such as 'repeatedly posting the same or similar comments', using link-farmed content and harmful third-party content, and using bots to promote content (Reddit, 2018c).

In the User Agreement, Reddit 'reserves the right to monitor' the platform and using the service means that there is an implied agreement to such monitoring, but adds, 'At the same time, we do not guarantee we will monitor at all' (Reddit, 2018c). Like all the other companies included here, users are responsible for content associated with their account and if the content violate policies, the content or user account can be removed.

Advertising Content

Reddit has Advertising Terms and an Advertising Policy. The terms require that the advertiser provides 'factual', 'true', and 'accurate' information – a much higher bar than for user content (Reddit, 2018b). The Advertising Policy reinforces the standard for advertising content, requiring advertisements not to 'employ techniques that are deceptive, untrue, or misleading' (Reddit, 2018a). The policy also directly prohibits advertisements that are 'designed to cause inflation of website metrics, such as clicks, impressions, likes, or followers' – in other words, 'clickbait' (Reddit, 2018a). Furthermore, Reddit stresses that it 'maintains high standards for not only the content and style of advertisements, but also the URL and landing page of any promoted website' (Reddit, 2018a). As some computational propaganda is used not for political but for monetary gains, the explicit terms against such use of advertisements may assist the platform in addressing clickbait and other forms of junk content. Finally, with reference to political advertising, the Advertising Policy prohibits 'intolerant or overly contentious political or cultural topics or views' (Reddit, 2018a). All other political advertising, including advertising relating to political issues, political committees, or candidates, or ads soliciting contributions, are 'subject to preapproval' and are required to comply with local laws and regulations.

Table 8. Summary of platforms' user content terms and policies

Term	Google	YouTube	Facebook	Twitter	LinkedIn	Amazon	Reddit
User has sole responsibility for content	X	X	X	X	X	X	X
User has right to opt out of curated content							
User has right to opt out of targeted advertising	X	X	No	X (via device and browser controls)	X	X (third party only)	X (third party only)
Policy banning user content that is illegal, misleading, shocking, spam, etc.	illegal	illegal, misleading, shocking, spam	illegal, misleading, inaccurate, impersonation	truth, illegal, unlawful, fake, misleading, confusing, deceive	illegal, shocking, intimidating, trick, fraud, defamatory, misleading, inaccurate	Illegal, misleading, inauthentic, defamatory, injurious, false (identity)	illegal, impersonation
Direct reference to false information, misinformation, or fake news					false information		
Relevant spam policy		X		X	X	X	X
Policies include right of provider to review content	X	X	X	X		X	X
Policies include specific right of provider to do automated review of content	X	X			(private messages)		
Policies include specific right of provider to do manual review of content		X					
Policies include right to take down material that is not illegal in provider's home territory	X	X	X	X	X	X	X
Right to take down content deemed threat to public safety (harm, violence, etc.)		X	X	X	X	X	X
Right to take down posts without notification	X	X	X	X	X	X	X
Right to suspend user accounts or services without notice	X	X	X	X	X	X	X
Right to limit visibility of content			X	X			

Table 9. Summary of platforms' advertising terms and policies

Term	Google	YouTube	Facebook	Twitter	LinkedIn	Amazon	Reddit
Has advertiser guidelines	AdWords, AdSense	AdWords, AdSense	Advertising policies	Ads policies, Ads MSA	Advertising guidelines	Creative Acceptance Policy	Advertising Policy, Advertising Terms
Direct reference to false/misinformation, fake news, etc.	false		false	false	false, lie	false	factual, true, untrue
Policy banning advertisements with illegal, misleading, shocking content, spam, etc.	dishonest, misleading, deceptive, misrepresentative, inaccurate	illegal, misleading, deceptive, shocking	misleading, sensational, shocking	deceptive, spam, inappropriate, authenticity, honest, misleading	fraudulent, deceptive, misleading, exaggerate	deceptive, misleading	accurate, deceptive, untrue, or misleading, non-deceptive, defensible
Language to address use of fake news in drawing ad revenues, likes, promotions, etc.							
Allows political advertising	X	X	X	X	X	X	X
Right to analyse content/data/profiles for tailored advertising	X	X	X	X	X	X	X
Advertisements reviewed/approved before posting	X	X	X	can be reviewed		X	X
Automated/machine review of advertisements							
Manual/human review of advertisements							
Right to terminate account of advertisement publishers	X	X	X	X	X		X
Right to take down ads without notification			X	X	X		X

7 CONCLUSION

After a slow start, there has been a raft of self-regulatory responses from several of the six platforms studied for this paper, especially Google and YouTube, Facebook, and Twitter. A key area for intervention is improved enforcement of existing terms and policies, as well as some steps towards increased collaboration with other actors, including news media, fact-checkers, and civil society organizations. We found little evidence of major changes to the underlying documents. This may be because of the limited versioning resources available for comparison, or because policies are currently being updated but have not yet been made public. It is expected that policies will be modified in the near future – either to take into account new legislation such as the GDPR or in response to government pressure following formal inquiries into fake news and scandals such as the Cambridge Analytica revelations.

The language used in the existing policies is broad enough to enable companies to apply the policies to a range of issues related to computational propaganda. Commonly used terms include illegal, unlawful, deceitful, and misleading. Specific sections on spam in the companies' policies were found to be particularly applicable to junk news and political bots, as it is often defined broadly and linked to repeated posting or sending of unwanted messages. Advertising policies included the most robust language relevant to elements of computational propaganda related to political campaigning, dark posts, and micro-targeting advertisements. Terms in advertising policies often linked content to language like deception, false, misleading, and truth.

For the Internet platforms studied for this report, 2016 was a defining moment. The ongoing shocks relating to election interference, junk news, and Cambridge Analytica, combined with deeper concerns about the viability of the business model of established news media and the impact of technology on democracy, all conspired to undermine citizens' and public authorities' confidence in platform companies. Initially, the platforms fell back on traditional postures – trying to minimize the perceived impact, for instance by quoting statistics highlighting how few accounts were involved – but the chronology of industry responses in this report tracks a changing attitude. However, there may be trouble ahead as Google and Facebook appear to be taking conflicting stances on their responsibility (or lack thereof) for content (Nicolaou & Bradshaw, 2018). As government regulation appears inevitable, the platforms have formulated numerous solutions to combat computational propaganda. Yet, despite 18 months of inquiries and bad press, there is little evidence of significant changes to company terms and policies which grant extensive powers over users' content, data, and behaviour. For meaningful progress to be made, and trust to be restored, there needs to be equitable relationships between the platforms and the people who use them.

8 APPENDICES

8.1 Appendix 1: Examples of Country-Specific Initiatives

Table 10. Examples of country-specific initiatives (other than USA)

Country	Platform	Description of Initiative
Austria	Google	New partners – fact-checking organizations from Brazil, Austria, and Spain (Anderson, 2017).
Brazil	Google	New partners – fact-checking organizations from Brazil, Austria, and Spain (Anderson, 2017).
	Google	Empowering underrepresented voices. Better diversity in the newsrooms. Partnered with organizations in the US, Brazil, France, and Germany. Expanding News Lab to Brazil and Asia training journalists from 15 countries (Grove, 2017b).
	Google	Providing support to partner First Draft Coalition in launch of Disinfo Lab – it employs journalists to leverage computational tools to monitor misinformation in run-up to elections – for 2018 elections in Mexico, Brazil, Indonesia, and US. Media literacy – supporting programmes in the UK, Brazil, and Canada (Gingras, 2018).
	Facebook	Elections will continue to be a major focus, e.g., India, Brazil, Mexico, Pakistan, Hungary, and others (Facebook Newsroom, 2018).
Canada	Google	Media literacy – supporting programmes in the UK, Brazil, and Canada (Gingras, 2018).
	Google	Google funds teaching Canadian students to spot fake news (Toronto Sun, 2017).
	Facebook	'View Ads' function and searchable political archive – Canada (Goldman, 2018a).
	Facebook	Starting December 2017, 'people will be able to click 'View Ads' on a page and it will run on Facebook, Instagram, and Messenger – whether or not the person viewing is part of the intended target audience for the ad. All pages will be part of this effort, and we will require that all ads be associated with a page as part of the ad creation process. We will start this test in Canada and roll it out to the US by this summer [2018]. Testing in one market allows us to learn the various ways an entire population uses the feature at a scale that allows us to learn and iterate. Starting in Canada was a natural choice as this effort aligns with our election integrity work already underway there' (Goldman, 2017a).
	Facebook	Canadian Election Integrity initiative: two-year partnership between Facebook Canada and MediaSmarts to provide digital and news literacy; release of Facebook Cyber Hygiene Guide for Politicians and Political Parties to support information security; launch of emergency Canadian Facebook Cyber Threats Crisis Email Line (to be used in the event of interference or suspected hacks); 'View Ads' (see above) (Facebook Canada, 2018).
	Facebook	Testing transparency across all ads in Canada (Facebook Newsroom, 2018a).
	Google, Facebook	Media Smarts, Canada's Centre for Digital and Media Literacy; Google and Facebook listed as sponsors (MediaSmarts, 2018).
France	Google	Empowering underrepresented voices. Better diversity in the newsrooms. Partnered with organizations in the US, Brazil, France, and Germany. In France, launched web app showing search interest in candidates (Grove, 2017b).
	Facebook	Changes to improve the ability of 'systems' to detect fake accounts. In France, these changes enabled Facebook to take action against over 30,000 fake accounts (Facebook Security, 2017).

Country	Platform	Description of Initiative
	Facebook	Fact-checking photos and videos in addition to links, starting in France with the AFP and 'scaling to more countries and partners soon' (Facebook Newsroom, 2018a).
Germany	Google	First Draft Coalition – collaborative reporting models to combat misinformation in UK, French, and German elections; better diversity in the newsrooms. Partnered with organizations in the US, Brazil, France, and Germany; helping newsrooms use Google Trends data to support elections coverage; built a hub in Germany (Grove, 2017b).
	Facebook	German elections: new playbook for working with local election commission (Facebook Newsroom, 2018a).
	Facebook	<i>Facebook's crackdown ahead of German election shows it's learning</i> (Sept. 2017); '...it shut down tens of thousands of fake accounts ahead of last weekend's [24 Sept. 2017] election in Germany...'. Other actions – tested related articles (giving different perspectives on articles), used machine learning to reduce clickbait and spam, provided a tab in the application where political parties could set out their position on key issues; launched an Election Hub experience; cooperated with German authorities and supported national and international media partners (Lapowsky, 2017).
	Facebook	'Partnering with public authorities like the Federal Office for Information Security (Germany)' (Zuckerberg, 2017).
	Facebook	Developed a machine-learning model that can detect different types of engagement bait. Posts identified will be shown less in news feeds. Roll-out over 'several weeks'. Update on 4 April 2018 – now expanded to Spanish, Arabic, French, Italian, and German posts (Silverman & Huang, 2017).
Hungary	Facebook	Elections will continue to be a major focus, e.g., India, Brazil, Mexico, Pakistan, Hungary, and others (Facebook Newsroom, 2018).
India	Google	New partners – fact-checking organizations from Brazil, Austria, and Spain. Plans to expand to Indonesia, Japan, and India (Anderson, 2017).
	Facebook	Elections will continue to be a major focus, e.g., India, Brazil, Mexico, Pakistan, Hungary, and others (Facebook Newsroom, 2018).
Indonesia	Google	New partners – fact-checking organizations from Brazil, Austria, and Spain. Plans to expand to Indonesia, Japan, and India (Anderson, 2017).
	Google	Providing support to partner First Draft Coalition in launch of Disinfo Lab – it employs journalists to leverage computational tools to monitor misinformation in run-up to elections – for 2018 elections in Mexico, Brazil, Indonesia, and US. (Gingras, 2018).
	Facebook	A machine-learning model that can detect different types of engagement. Identified bait posts will be shown less in newsfeeds' roll-out over 'several weeks'. Update on 26 April 2018 – now expanded to posts in Mandarin, Hindi, Indonesian, Tagalog, Vietnamese, Thai, and Turkish (Silverman & Huang, 2017).
Italy	Facebook	Fact-checking partners in six countries, including Italy and Mexico (Facebook Newsroom, 2018a).
Mexico	Google	Providing support to partner First Draft Coalition in launch of Disinfo Lab – it employs journalists to leverage computational tools to monitor misinformation in run-up to elections – for 2018 elections in Mexico, Brazil, Indonesia, and US. Media literacy – supporting programmes in the UK, Brazil, and Canada (Gingras, 2018).
	Facebook	Elections will continue to be a major focus, e.g., India, Brazil, Mexico, Pakistan, Hungary, and others (Facebook Newsroom, 2018).
	Facebook	Fact-checking partners in six countries, including Italy and Mexico (Facebook Newsroom, 2018a).

Country	Platform	Description of Initiative
	Facebook	Facebook signed a 'Memorandum of Cooperation' with the National Electoral Institute (INE) of Mexico to support the upcoming July [2018] elections. The INE is committed to providing election data to Facebook; 'Facebook intends (but does not commit) to make certain of its civic engagement products available on its platform to its users in Mexico' (Tadeo & Dunne, 2018).
Pakistan	Facebook	Elections will continue to be a major focus, e.g., India, Brazil, Mexico, Pakistan, Hungary, and others (Facebook Newsroom, 2018d).
Spain	Google	New partners – fact-checking organizations from Brazil, Austria, and Spain (Anderson, 2017).
United Kingdom	Google	Launch of Internet Citizens in Liverpool, UK, a media literacy programme empowering users to combat echo chambers, identify fake news, flag content, and increase their sense of confidence online. Partnership with third parties (Institute for Strategic Dialogue, UK Youth, and Livity). Use of video (YouTube) to combat social division (Grummer, 2017).
	Google	Combating misinformation in European elections. First Draft Coalition – collaborative reporting models used to combat misinformation in UK, French, and German elections (Grove, 2017b).

8.2 Appendix 2: Terms of Service Analysed

Table 11. LinkedIn – Terms of service

Document	Additional Versions Available	Current & Add'l Versions Since November 2016
Service Terms	No	
User Agreement	Yes (one)	7 June 2017 effective 8 May 2018
Professional Community Guidelines Professional Community Policies	Yes (one)	7 June 2016 effective 8 May 2018
Privacy Policy	Yes (one)	7 June 2017 effective 8 May 2018
Advertising Guidelines	No	13 December 2017
Publishing Platform Guidelines	No	June 2016

Table 12. Reddit – Terms of service

Document	Additional Versions Available	Current & Add'l Versions Since November 2016
Reddit User Agreement and Reddit Rules	Yes	21 March 2018 7 December 2017 27 September 2017 19 June 2017
Reddit Content Policy	No	
Reddit, Inc. Privacy Policy	Yes	12 December 2017 21 August 2017 22 November 2016
Advertising Terms: Ads Platform	No	19 January 2018
Advertising Policy	No	10 November 2017
Moderator Guidelines for Healthy Communities	No	17 April 2017

Table 13. Twitter – Terms of service

Document	Additional Versions Available	Current & Add'l Versions Since November 2016
Twitter Terms of Service (living outside the US)	Yes	2 October 2017
The Twitter Rules	No	
Twitter Privacy Policy	Yes	18 June 2017
Twitter Ads Policies	Yes	30 August 2017
Twitter Ads Master Service Agreement (MSA)	No	

Table 14. Google – Terms of service

Document	Additional Versions Available	Current & Add'l Versions Since November 2016
Google Terms of Service	Yes	25 October 2017 18 December 2017
Privacy Policy	Yes	2 October 2017 17 April 2017 1 March 2017
Advertising – Privacy & Terms	No	
AdWords Advertising Policies	Yes (by topic and country)	November 2016

Table 15. YouTube – Terms of service

Document	Additional Versions Available	Current & Add'l Versions Since November 2016
Terms of Service (YouTube)	No	9 June 2010
Community Guidelines (YouTube)	No	
YouTube Policies	No	
Google Privacy Policy	Yes	18 December 2017 2 October 2017 17 April 2017 1 March 2017
Advertiser Friendly Content Guidelines (YouTube)	No	
Google AdSense Program Policies	No	

Table 16. Facebook – Terms of service

Document	Additional Versions Available	Current & Add'l Versions Since November 2016
Statement of Rights and Responsibilities	No	31 January 2018
Facebook Principles	No	
Community Standards	No	
Data Policy	No	29 September 2016
Advertising Policies	No	

9 REFERENCES

- Albrecht, J. (2018, March 20). Introducing Subscribe with Google. Retrieved from <https://www.blog.google/topics/google-news-initiative/introducing-subscribe-google/>
- Anderson, E. (2017, July 6). Making it easier for publishers to share fact check content. Retrieved from <https://www.blog.google/topics/journalism-news/making-it-easier-publishers-share-fact-check-content/>
- Anker, A., Su, S., & Smith, J. (2017, October 5). New test to provide context about articles. Retrieved from <https://newsroom.fb.com/news/2017/10/news-feed-fyi-new-test-to-provide-context-about-articles/>
- Bala, D. (2017, December 19). Targeting exclusions - Update. Retrieved from <https://developers.facebook.com/ads/blog/post/2017/12/19/targeting-exclusions-update-blog-post>
- BBC. (2015, December 3). Hackers befriend LinkedIn users. *BBC News*. Retrieved from <http://www.bbc.com/news/technology-34994858>
- Bickert, M. (2018a, April 24). Hard questions: Why do you leave up some posts but take down others? Retrieved from <https://newsroom.fb.com/news/2018/04/community-standards-examples/>
- Bickert, M. (2018b, April 24). Publishing our internal enforcement guidelines and expanding our appeals process. Retrieved from <https://newsroom.fb.com/news/2018/04/comprehensive-community-standards/>
- Bradshaw, S., & Neudert, L.-M. (2018). Government responses to social media manipulation. *Computational Propaganda Project Working Paper*.
- Cambridge English Dictionary. (2018). Fake news [definition]. Retrieved from <https://dictionary.cambridge.org/dictionary/english/fake-news>
- Cevahir, H. K. (2017, April 17). Breaking new ground in the fight against fake Facebook likes. Retrieved from <https://www.facebook.com/business/news/fighting-fake-likes-update>
- Chakrabarti, S. (2018, January 22). Hard questions: What effect does social media have on democracy? Retrieved from <https://newsroom.fb.com/news/2018/01/effect-social-media-democracy/>
- Chang, J. (2017, November 16). Identifying credible content online, with help from the Trust Project. Retrieved from <https://www.blog.google/topics/journalism-news/sorting-through-information-help-trust-project/>

- Chen, A. (2014, October 23). The laborers who keep dick pics and beheadings out of your Facebook feed. *Wired*. Retrieved from <https://www.wired.com/2014/10/content-moderation/>
- Cohen, T. (2017, March 22). Introducing trending storylines: Discover trending news and views on LinkedIn. Retrieved from <https://blog.linkedin.com/2017/march/22/introducing-trending-storylines-discover-trending-news-and-views-on-LinkedIn>
- Collins, B., & Russell, J. (2018, January 3). Russians used Reddit and Tumblr to troll the 2016 election. Retrieved from <https://www.thedailybeast.com/russians-used-reddit-and-tumblr-to-troll-the-2016-election>
- Confessore, N., Dance, G. J. X., & Harris, R. (2018, January 31). Twitter followers vanish amid inquiries into fake accounts. *The New York Times*. Retrieved from <https://www.nytimes.com/interactive/2018/01/31/technology/social-media-bots-investigations.html>
- Crowel, C. (2017, June 14). Our approach to bots & misinformation. Retrieved from https://blog.twitter.com/official/en_us/topics/company/2017/Our-Approach-Bots-Misinformation.html
- Dieudonne, D. (2017, November). Fact-checking the French election: Lessons from CrossCheck, a collaborative effort to combat misinformation. Retrieved from <https://www.blog.google/topics/journalism-news/fact-checking-french-election-lessons-crosscheck-collaborative-effort-combat-misinformation/>
- DoubleClick. (2017, October 3). Building for beyond with the Insights Engine Project. Retrieved from <https://doubleclick-publishers.googleblog.com/2017/10/building-for-beyond-with-insights.html>
- Edgett, S. (2017, October 31). U.S. Senate Committee on the Judiciary: Opening remarks. Retrieved from https://blog.twitter.com/official/en_us/topics/company/2017/opening_remarks.html
- Everson, C. (2017, September 13). Providing more clarity and controls for advertisers. Retrieved from <https://www.facebook.com/business/news/providing-more-clarity-and-controls-for-advertisers>
- Facebook. (2017, April 13). Improvements in protecting the integrity of activity on Facebook. Retrieved from <https://www.facebook.com/notes/facebook-security/improvements-in-protecting-the-integrity-of-activity-on-facebook/10154323366590766/>

- Facebook. (2018a). Advertising Policies. Retrieved May 6, 2018, from <https://www.facebook.com/business/help/223106797811279>
- Facebook. (2018b). Community Standards. Retrieved from <https://www.facebook.com/communitystandards/>
- Facebook. (2018c). Terms of Service - Statement of Rights and Responsibilities. Retrieved from <https://www.facebook.com/terms.php>
- Facebook Business. (2017, September 15). Updating our ad targeting. Retrieved from <https://www.facebook.com/business/news/updating-our-ad-targeting>
- Facebook Canada. (2018). Facebook Canadian Election Integrity Initiative. Retrieved May 7, 2018, from <http://facebookcanadianelectionintegrityinitiative.com/>
- Facebook Newsroom. (2017a, August 8). Updates to Clicks and Impression Reporting. Retrieved from <https://newsroom.fb.com/news/2017/08/updates-to-clicks-and-impression-reporting/>
- Facebook Newsroom. (2017b, November 22). Continuing transparency on Russian activity. Retrieved from <https://newsroom.fb.com/news/2017/11/continuing-transparency-on-russian-activity/>
- Facebook Newsroom. (2018a, March 19). Hard questions: What is Facebook doing to protect election security? Retrieved from <https://newsroom.fb.com/news/2018/03/hard-questions-election-security/>
- Facebook Newsroom. (2018b, March 19). Pursuing forensic audits to investigate Cambridge Analytica claims. Retrieved from <https://newsroom.fb.com/news/2018/03/forensic-audits-cambridge-analytica/>
- Facebook Newsroom. (2018c, March 21). Cracking down on platform abuse. Retrieved from <https://newsroom.fb.com/news/2018/03/cracking-down-on-platform-abuse/>
- Facebook Newsroom. (2018d, April 4). Hard questions: Q&A with Mark Zuckerberg on protecting people's information. Retrieved from <https://newsroom.fb.com/news/2018/04/hard-questions-protecting-peoples-information/>
- Facebook Security. (2017, April 13). Improvements in protecting the integrity of activity on Facebook. Retrieved from <https://www.facebook.com/notes/facebook-security/improvements-in-protecting-the-integrity-of-activity-on-facebook/10154323366590766/>
- Falck, B. (2017, October 24). New transparency for ads on Twitter. Retrieved from https://blog.twitter.com/official/en_us/topics/product/2017/New-Transparency-For-Ads-on-Twitter.html

- Gingras, R. (2018, March 20). Elevating quality journalism on the open web. Retrieved from <https://www.blog.google/topics/google-news-initiative/elevating-quality-journalism/>
- Goldman, R. (2017a, October 27). Update on our advertising transparency and authenticity efforts. Retrieved from <https://newsroom.fb.com/news/2017/10/update-on-our-advertising-transparency-and-authenticity-efforts/>
- Goldman, R. (2017b, November 27). Our advertising principles. Retrieved from <https://newsroom.fb.com/news/2017/11/our-advertising-principles/>
- Goldman, R. (2018a, April 6). Making ads and pages more transparent. Retrieved from <https://newsroom.fb.com/news/2018/04/transparent-ads-and-pages/>
- Goldman, R. (2018b, April 23). Hard questions: What information do Facebook advertisers know about me? Retrieved from <https://newsroom.fb.com/news/2018/04/data-and-advertising/>
- Gomes, B. (2017, April 25). Our latest quality improvements for Search. Retrieved from <https://blog.google/products/search/our-latest-quality-improvements-search/>
- Google. (2018a). AdSense Program Policies. Retrieved May 6, 2018, from <https://support.google.com/adsense/answer/48182?hl=en>
- Google. (2018b). Advertising. Retrieved from <https://policies.google.com/technologies/ads>
- Google. (2018c). AdWords Policies. Retrieved from <https://support.google.com/adwordspolicy/answer/6008942?hl=en>
- Google. (2018d). Be Internet Awesome - About. Retrieved from <https://beinternetawesome.withgoogle.com/en/>
- Google. (2018e). Google Terms of Service. Retrieved from <https://policies.google.com/terms>
- Greenleaf, K. (2017, June 27). Developing critical reading skills with media literacy apps on Chromebooks. Retrieved from <https://www.blog.google/topics/education/developing-critical-reading-skills-media-literacy-apps-chromebooks/>
- Grewal, P. (2018, March 16). Suspending Cambridge Analytica and SCL Group from Facebook. Retrieved from <https://newsroom.fb.com/news/2018/03/suspending-cambridge-analytica/>

- Grove, S. (2017a, December 12). News Lab in 2017: Working with news organizations to address industry challenges. Retrieved from <https://www.blog.google/topics/journalism-news/news-lab-2017-working-news-organizations-address-industry-challenges/>
- Grove, S. (2017b, December 15). News Lab in 2017: The year in review. Retrieved from <https://www.blog.google/topics/journalism-news/news-lab-year-in-review/>
- Grummer, N. (2017, April 21). Internet citizens: Let's make a better web. Retrieved from <https://www.blog.google/topics/google-europe/migration-folder-googlepolicyeuropeblogspotcom-2016-08-04-185442/internet-citizens-lets-make-better-web/>
- Hamodi, B., Bokhari, Z., & Zhang, Y. (2017, August 17). Taking action against video clickbait. Retrieved from <https://newsroom.fb.com/news/2017/08/news-feed-fyi-taking-action-against-video-clickbait/>
- Harrington, S. (2018, March 8). Updates to our Terms of Service. Retrieved from <https://blog.linkedin.com/2018/march/8/updates-to-our-terms-of-service>
- Harvey, D. (2018, April 5). Serving the public conversation during breaking events. Retrieved from https://blog.twitter.com/official/en_us/topics/company/2018/Serving-the-Public-Conversation-During-Breaking-Events.html
- Huffmann, S. (2018, March 5). In response to recent reports about the integrity of Reddit, I'd like to share our thinking. Retrieved from https://www.reddit.com/r/announcements/comments/827zqc/in_response_to_recent_reports_about_the_integrity/
- Johnson, E. (2016, December 29). Why is there no fake news on LinkedIn? Listen to its editor explain. Retrieved from <https://www.recode.net/2016/12/29/14100064/linkedin-daniel-roth-fake-news-facebook-recode-podcast>
- Kaplan, J. (2017, October 2). Improving enforcement and transparency of ads on Facebook. Retrieved from <https://newsroom.fb.com/news/2017/10/improving-enforcement-and-transparency/>
- Kosslyn, J., & Yu, C. (2017, April 7). Fact Check now available in Google Search and News around the world. Retrieved from <https://www.blog.google/products/search/fact-check-now-available-google-search-and-news-around-world/>
- Lapowsky, I. (2017, September 27). Facebook's crackdown ahead of German election shows it's learning. Retrieved from <https://www.wired.com/story/facebooks-crackdown-ahead-of-german-election->

shows-its-learning; <https://de.newsroom.fb.com/news/2017/09/update-zu-den-wahlen/>

Lapowsky, I. (2018, March 9). Reddit still hosts links to Russian propaganda sites. Retrieved from <https://www.wired.com/story/reddit-russian-propaganda/>

Levin, S., Wong, J. C., & Harding, L. (2016, September 9). Facebook backs down from 'napalm girl' censorship and reinstates photo. *The Guardian*. Retrieved from <http://www.theguardian.com/technology/2016/sep/09/facebook-reinstates-napalm-girl-photo>

LinkedIn. (2018a). LinkedIn Advertising Guidelines. Retrieved May 6, 2018, from <https://www.linkedin.com/legal/sas-guidelines>

LinkedIn. (2018b). LinkedIn Professional Community Policies. Retrieved May 6, 2018, from <https://www.linkedin.com/help/linkedin/answer/89880/linkedin-professional-community-policies?lang=en>

LinkedIn. (2018c). Privacy Policy. Retrieved May 6, 2018, from <https://www.linkedin.com/legal/privacy-policy>

LinkedIn. (2018d). User Agreement. Retrieved from <https://www.linkedin.com/legal/user-agreement>

Lyons, T. (2017, December 20). Replacing disputed flags with related articles. Retrieved from <https://newsroom.fb.com/news/2017/12/news-feed-fyi-updates-in-our-fight-against-misinformation/>

Marantz, A. (2018, March 19). Reddit and the struggle to detoxify the Internet. Retrieved from <https://www.newyorker.com/magazine/2018/03/19/reddit-and-the-struggle-to-detoxify-the-internet>

MediaSmarts. (2018). Sponsors. Retrieved from <https://mediasmarts.ca/about-us/sponsors>

Mosseri, A. (2016, December 15). Addressing hoaxes and fake news. Retrieved from <https://newsroom.fb.com/news/2016/12/news-feed-fyi-addressing-hoaxes-and-fake-news/>

Mosseri, A. (2017a, April 6). A new educational tool against misinformation. Retrieved from <https://newsroom.fb.com/news/2017/04/a-new-educational-tool-against-misinformation/>

Mosseri, A. (2017b, April 6). Working to stop misinformation and false news. Retrieved from <https://newsroom.fb.com/news/2017/04/working-to-stop-misinformation-and-false-news/>

- Mosseri, A. (2018, January 19). Helping ensure news on Facebook is from trusted sources [Facebook Newsroom]. Retrieved from <https://newsroom.fb.com/news/2018/01/trusted-sources/>
- Nicolaou, A., & Bradshaw, T. (2018, March 22). Facebook and Google executives clash over 'fake news.' Retrieved from <https://www.ft.com/content/3929bb2c-2de9-11e8-9b4b-bc4b9f08f381>
- Oltermann, P. (2018, January 5). Tough new German law puts tech firms and free speech in spotlight. *The Guardian*. Retrieved from <https://www.theguardian.com/world/2018/jan/05/tough-new-german-law-puts-tech-firms-and-free-speech-in-spotlight>
- Pasquale, F. (2017, February 6). From Holocaust denial to Hitler admiration, Google's algorithm is dangerous. Retrieved from https://www.huffingtonpost.com/entry/holocaust-google-algorithm_us_587e8628e4b0c147f0bb9893
- Reddit. (2017, December 12). Privacy Policy. Retrieved from <https://www.redditinc.com/policies/privacy-policy>
- Reddit. (2018a). Advertising on Reddit. Retrieved from <https://www.redditinc.com/advertising>
- Reddit. (2018b). Advertising Terms: Ads Platform. Retrieved May 6, 2018, from <https://www.reddithelp.com/en/categories/advertising/policy-and-guidelines/advertising-terms-ads-platform>
- Reddit. (2018c). Content Policy. Retrieved from <https://www.redditinc.com/policies/content-policy>
- Reddit. (2018d). Moderator guidelines for healthy communities. Retrieved from <https://www.reddit.com/help/healthycommunities/>
- Reddit. (2018e). Reddit's 2017 transparency report and suspect account findings. Retrieved May 7, 2018, from https://www.reddit.com/r/announcements/comments/8bb85p/reddits_2017_transparency_report_and_suspect/
- Reddit. (2018f, March 21). User Agreement. Retrieved from <https://www.redditinc.com/policies/user-agreement>
- Roberts, S. T. (2016, May 2). Behind the screen: The people and politics of commercial content moderation. Retrieved from <http://opentranscripts.org/transcript/politics-of-commercial-content-moderation/>

- Romm, T. (2018, March 5). Senate investigators want answers from Reddit and Tumblr on Russia meddling. *The Washington Post*. Retrieved from https://www.washingtonpost.com/news/the-switch/wp/2018/03/05/senate-investigators-want-answers-from-reddit-and-tumblr-on-russia-meddling/?utm_term=.9617ae037259
- Rosen, G. (2018, March 29). Hard questions: What is Facebook doing to protect election security? Retrieved from <https://newsroom.fb.com/news/2018/03/hard-questions-election-security/>
- Roth, Y. (2018, February 21). Automation and the use of multiple accounts. Retrieved from https://blog.twitter.com/developer/en_us/topics/tips/2018/automation-and-the-use-of-multiple-accounts.html
- Rushe, D. (2018, April 9). Zuckerberg's testimony: CEO will defend Facebook as 'positive force'. *The Guardian*. Retrieved from <https://www.theguardian.com/us-news/2018/apr/09/mark-zuckerberg-facebook-testimony-congress>
- Schindler, P. (2018, March 20). The Google News Initiative: Building a stronger future for news. Retrieved from <https://www.blog.google/topics/google-news-initiative/announcing-google-news-initiative/>
- Schrage, E., & Ginsberg, D. (2018, April 9). Facebook launches new initiative to help scholars assess social media's impact on elections [Facebook Newsroom]. Retrieved from <https://newsroom.fb.com/news/2018/04/new-elections-initiative/>
- Schroepfer, M. (2018, April 4). An update on our plans to restrict data access on Facebook [Facebook Newsroom]. Retrieved from <https://newsroom.fb.com/news/2018/04/restricting-data-access/>
- Shukla, S., & Lyons, T. (2017, August 28). Blocking ads from pages that repeatedly share false news. Retrieved from <https://newsroom.fb.com/news/2017/08/blocking-ads-from-pages-that-repeatedly-share-false-news/>
- Silverman, H., & Huang, L. (2017, December 18). Fighting engagement bait on Facebook. Retrieved from <https://newsroom.fb.com/news/2017/12/news-feed-fyi-fighting-engagement-bait-on-facebook/>
- Simo, F. (2017, January 11). Announcing the Facebook Journalism Project. Retrieved from <https://newsroom.fb.com/news/2017/01/announcing-the-facebook-journalism-project/>

- Stamos, A. (2017, September 6). An update on information operations on Facebook. Retrieved from <https://newsroom.fb.com/news/2017/09/information-operations-update/>
- Stamos, A. (2018, April 3). Authenticity matters: The IRA has no place on Facebook. Retrieved from <https://newsroom.fb.com/news/2018/04/authenticity-matters/>
- Stretch, C. (2017, September 21). Facebook to provide Congress with ads linked to Internet research agency. Retrieved from <https://newsroom.fb.com/news/2017/09/providing-congress-with-ads-linked-to-internet-research-agency/>
- Tadeo, J., & Dunne, L. (2018, February). Mexico's electoral authority announces collaboration with Facebook, casting doubt on its credibility. Retrieved from <https://advox.globalvoices.org/2018/02/23/mexicos-electoral-authority-announces-collaboration-with-facebook-casting-doubt-on-its-credibility/>
- Taylor, E. (2016). *The privatization of human rights: Illusions of consent, automation and neutrality*. Waterloo, Ontario: Centre for International Governance Innovation & London: Chatham House. Retrieved from https://ourinternet-files.s3.amazonaws.com/publications/no24_web_2.pdf
- Toronto Sun. (2017, September 19). Google funds teaching Canadian students to spot fake news. Retrieved from <http://torontosun.com/2017/09/19/google-funds-teaching-canadian-students-to-spot-fake-news/wcm/e31d39fa-ba94-4941-90d1-6b0374cccdb8>
- Tufekci, Z. (2018, March 11). YouTube, the great radicalizer [Opinion]. *The New York Times*. Retrieved from <https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html>
- Twitter. (2017, October). Twitter Terms of Service. Retrieved from <https://twitter.com/content/twitter-com/legal/en/tos.html>
- Twitter. (2018a). Ad Policy: Inappropriate content. Retrieved May 6, 2018, from <https://business.twitter.com/en/help/ads-policies/prohibited-content-policies/inappropriate-content.html>
- Twitter. (2018b). Advertising potentially offensive content policy and best practices. Retrieved from <https://business.twitter.com/en/help/ads-policies/other-policy-requirements/offensive-content-policy-and-best-practices.html>
- Twitter. (2018c). Privacy Policy. Retrieved from <https://twitter.com/content/twitter-com/legal/en/privacy.html>
- Twitter. (2018d). The Twitter rules. Retrieved May 6, 2018, from <https://help.twitter.com/en/rules-and-policies/twitter-rules>

- Twitter. (2018e). Twitter Ads Policies. Retrieved from <https://business.twitter.com/en/help/ads-policies/introduction-to-twitter-ads/twitter-ads-policies.html>
- Twitter. (2018f). Twitter Master Services Agreement. Retrieved from <https://legal.twitter.com/ads-terms/us.html>
- Twitter Public Policy. (2017a, September 28). Update: Russian interference in 2016 US election, bots, & misinformation. Retrieved from https://blog.twitter.com/official/en_us/topics/company/2017/Update-Russian-Interference-in-2016--Election-Bots-and-Misinformation.html
- Twitter Public Policy. (2017b, October 26). Announcement: RT and Sputnik advertising. Retrieved from https://blog.twitter.com/official/en_us/topics/company/2017/Announcement-RT-and-Sputnik-Advertising.html
- Twitter Public Policy. (2018, January 19). Update on Twitter's Review of the 2016 U.S. Election. Retrieved from https://blog.twitter.com/official/en_us/topics/company/2018/2016-election-update.html
- Walker, K., & Salgado, R. (2017, October 30). Security and disinformation in the U.S. 2016 election. Retrieved from <https://www.blog.google/topics/public-policy/security-and-disinformation-us-2016-election/>
- Washington Post. (2018). Transcript of Mark Zuckerberg's Senate hearing. Retrieved from https://www.washingtonpost.com/news/the-switch/wp/2018/04/10/transcript-of-mark-zuckerbergs-senate-hearing/?utm_term=.307b8afb1714
- Weedon, J., Nuland, W., & Stamos, A. (2017). *Information operations and Facebook. Facebook Security*. Retrieved from <https://fbnewsroomus.files.wordpress.com/2017/04/facebook-and-information-operations-v1.pdf>
- Woolley, S. C., & Howard, P. N. (2016). Automation, algorithms, and politics: Political communication, computational propaganda, and autonomous agents – Introduction. *International Journal of Communication*, 10, 9.
- YouTube. (2017a, December 4). Expanding our work against abuse of our platform. Retrieved from <https://youtube.googleblog.com/2017/12/expanding-our-work-against-abuse-of-our.html>
- YouTube. (2017b, December 4). Official YouTube blog: Expanding our work against abuse of our platform. Retrieved from

<https://youtube.googleblog.com/2017/12/expanding-our-work-against-abuse-of-our.html>

YouTube. (2018a). Advertiser-friendly content guidelines. Retrieved from <https://support.google.com/youtube/answer/6162278?hl=en>

YouTube. (2018b). Policies and safety. Retrieved May 6, 2018, from <https://www.youtube.com/yt/about/policies/#community-guidelines>

YouTube. (2018c, February). Greater transparency for users around news broadcasters. Retrieved from <https://youtube.googleblog.com/2018/02/greater-transparency-for-users-around.html>

YouTube. (2018d, February 2). Official YouTube blog: Greater transparency for users around news broadcasters. Retrieved from <https://youtube.googleblog.com/2018/02/greater-transparency-for-users-around.html>

Zuckerberg, M. (2016, November 19). Mark Zuckerberg – A lot of you have asked what we're doing about... Retrieved from <https://www.facebook.com/zuck/posts/10103269806149061>

Zuckerberg, M. (2017, September 21). Mark Zuckerberg – I just went live a minute ago, here's what I said. Retrieved from <https://www.facebook.com/zuck/posts/10104052907253171>

10 SERIES ACKNOWLEDGEMENT

The authors gratefully acknowledge the support of the European Research Council, Computational Propaganda: Investigating the Impact of Algorithms and Bots on Political Discourse in Europe, Proposal 648311, 2015-2020, Philip N Howard, Principal Investigator. Additional support has been provided by the Ford Foundation. Project activities were approved by the University of Oxford's Research Ethics Committee (CUREC OII C1A15-044). Any opinions, findings and conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funders or the University of Oxford.