

Artificial intelligence and the rights to assembly and association

Cameran Ashraf 

School of Public Policy, Central European University, Budapest, Hungary

ABSTRACT

The rights to assembly and association are fundamental rights guaranteed in the Universal Declaration of Human Rights. They are essential to the establishment and functioning of a democracy and ensure that individuals and groups can peacefully come together to pursue their common goals. These, and other human rights, are being challenged by the development and widespread deployment of artificial intelligence systems on the internet. Indeed, the development of artificial intelligence has been a cause for concern among human rights activists, scholars and practitioners.

While much existing literature has examined how AI will impact privacy and freedom of expression, its impact on the rights to assembly and association has been neglected. To develop a more well-rounded body of literature about AI and human rights, this paper will examine the impacts of artificial intelligence on the rights to assembly and association. It will discuss AI's impact on two key areas: content display, whereby AI determines the content we *see*, and content moderation, where AI determines which content *exists*. The paper concludes with policy recommendations and the hope that these recommendations will serve as a starting point for a discussion on protecting these important rights in the age of artificial intelligence.

ARTICLE HISTORY

Received 31 May 2019
Revised 7 February 2020
Accepted 13 March 2020


KEYWORDS

Artificial intelligence; human rights; algorithms; right to assembly and association

Introduction

Artificial intelligence (AI) is changing the world in dramatic ways. It manages billions of dollars with automated financial advisors, powers self-driving cars, determines credit-worthiness, assists law enforcement in deciding where and when to patrol and helps pilots fly. In the near future, AI will help doctors diagnose our ailments, grade university exams and mitigate biases in hiring practices (Raso et al. 2018).

On the internet, AI is used to moderate the content we see in our newsfeeds. In addition to independently removing content without human oversight, AI also acts as an aid to human reviewers in eliminating offensive content from online platforms. AI powers much of the internet we experience. It curates music and video playlists, provides recommendations for product purchases and scans comment sections for offensive messages.

CONTACT Cameran Ashraf  AshrafC@spp.ceu.edu

© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The rising public visibility of AI has encouraged vigorous debate in academia, policy circles and in the popular press. Much of this discussion has focused on the potential impact of AI on society and how to best regulate it. As a result, Google, Microsoft and other large tech companies have embraced AI ethics panels to address privacy and security concerns (Vincent 2019). Governments around the world have launched AI initiatives seeking competitive economic advantage, with China in particular seeking to be the global leader in AI by 2030 (Webster et al. 2017). In the domain of cyberwar, states are working to 'weaponize' AI for strategic advantage. These more popular debates have largely focused on AI and economics, ethics and conflict.

Discussion about AI's impact on human rights besides freedom of expression and privacy has been scarce (Raso et al. 2018). This is problematic, since internet is the de facto global commons. Every day, billions of people come together on the internet to discuss politics, daily life, illnesses, joys and sorrows. Religious groups discuss scripture and activists organise protests and air grievances. In other words, every day, we use the internet to exercise our freedom of assembly and association (FoAA), enumerated in Article 20 of the Universal Declaration of Human Rights.

Assembly and association are two separate human rights commonly grouped together. Freedom of assembly refers to 'the intentional and temporary presence of a number of individuals in a public place for a common expressive purpose' (Belyaeva et al. 2010, 15). Freedom of association 'enables individuals to come together and collectively to express, promote, pursue, and even defend common interests' (Smith and Anker 2005, 18). These two rights are 'essential both to establish a genuine democracy and to ensure that, once achieved, it remains healthy and flourishing' (Smith and Anker 2005, 18). Indeed, the United Nations Special Rapporteur on the rights to freedom of peaceful assembly and of association states: 'The rights to freedom of peaceful assembly and of association are key to the realisation of democracy and dignity, since they enable people to voice and represent their interests, to hold Governments to account and to empower human agency' (Voule 2018, 17). The rights to assembly and association are key not only to democracy, but also to the functioning of an open and free internet where billions of individuals come together.

Given the importance of AI for social media and the broader internet (VivaTech 2018), assembly and association online will be dependent upon how AI is used. By determining or influencing what content we see or what content exists, AI can shape how and why people assemble online, allowing some groups to exercise their rights while discouraging or banning others. Indeed, the former UN Special Rapporteur on the rights to freedom of peaceful assembly and of association, Maina Kiai, affirms that assembly and association exist online when he states that '... the rights to freedom of peaceful assembly and of association can be exercised through new technologies, including through the Internet' (Kiai 2012, 20).

There is a critical lack of literature on AI's impact on human rights besides freedom of expression and privacy (Raso et al. 2018). Therefore, this paper will provide an overview of AI and the rights to assembly and association online. It will discuss AI's impact on two key areas on the internet: content display, whereby AI determines the content we *see*, and content moderation, through which AI determines which content *exists*. Finally, it will provide preliminary policy recommendations. It is hoped that this paper will contribute to a broader human rights-centric discussion of AI and its current and future impacts.

What is AI?

AI has no single accepted definition. It is a broad field focused on enhancing the ability of computers to make ‘appropriate generalizations in a timely fashion based on limited data’ (Feldstein 2019a, 41) and is divided into two broad categories: narrow and general artificial intelligence.

General artificial intelligence seeks to develop systems which can learn new skills and convincingly imitate human behaviour (Privacy International and Article 19 2018). An example would be the AI assistant Samantha in the 2013 movie *Her*. Samantha was able to hold sophisticated conversations with Theodore, the protagonist. She understood how he felt, had opinions of her own and was capable of developing feelings. To date, no systems meet this definition and many experts in the field agree it will be feasible in 2099 at the earliest (Vincent 2018).

Narrow artificial intelligence’ is what we are familiar with; prominent examples include shopping recommendations, online translation services, automatically generated music playlists and customised social media feeds. Other examples include Google’s AlphaGo which beat Ke Jie, the world’s top Go player, or Google’s anti-suicide intervention when search terms suggest suicidal behaviour. Narrow AI works in narrow and restricted domains and seeks to imitate human talent. Here, the AI system attempts to ‘derive a rule or procedure that explains the data or can predict future data (White House 2016, 8).

Narrow AI is implemented through machine learning and algorithms. Generally, machine learning uses algorithms to analyse large amounts of data for a specific task. The Royal Society states that ‘Machine learning systems are set a task and given a large amount of data to use as examples of how this task can be achieved or from which to detect patterns. The system then learns how best to achieve the desired output’ (Royal Society 2017, 19).

These machine learning systems are powered by algorithms. An algorithm is a set of instructions telling a computer how to process data. They are ‘... computer code designed and written by humans, carrying instructions to translate data into conclusions, information or outputs’ (Kaye 2018, 4). Importantly, the algorithms which power machine learning are developed by computer engineers and programmers and therefore reflect their conscious and unconscious biases resulting from gender identity, socioeconomic class, culture and upbringing (Noble 2018). When the outputs of machine learning systems are used to make decisions, it is called ‘algorithmic decision making’.

The right to assembly and association and the internet – what it is and how it manifests

According to the Universal Declaration of Human Rights, everyone has the right to meet, engage in peaceful protest, collectively share their opinions, work, art and more (United Nations 1948). This right is a cornerstone of democracies and a means by which citizens hold governments and the powerful accountable. Broadly speaking, on the internet we exercise this by using social media, posting on forums, reading comments on news websites and countless other activities where we come together online.

The right to assembly and association is articulated in UDHR article 20, and states:

(1) ‘Everyone has the right to freedom of peaceful assembly and association.’

This is echoed in articles 21 and 22 of the *International Covenant on Civil and Political Rights* (ICCPR), ratified by 172 countries:

Article 21: The right of peaceful assembly shall be recognized. No restrictions may be placed on the exercise of this right other than those imposed in conformity with the law and which are necessary in a democratic society in the interests of national security or public safety, public order (*ordre public*), the protection of public health or morals or the protection of the rights and freedoms of others.

Article 22: 1. Everyone shall have the right to freedom of association with others, including the right to form and join trade unions for the protection of his interests.

Assembly and association are further enumerated in the European and American Convention on Human Rights, and the constitutions of the United States, Japan, Italy, Ireland, Spain, Turkey and elsewhere. It is a right which forms the basis for other human rights and is essential to the functioning of democracy (Voule 2018). Maina Kiai states that freedom of assembly and association empowers individuals to '... express their political opinions, engage in literary and artistic pursuits and other cultural, economic and social activities, engage in religious observances or other beliefs, form and join trade unions and cooperatives, and elect leaders to represent their interests and hold them accountable' (Kiai 2012, 5).

Freedom of assembly and association is well-established in human rights law and practice. However, the internet has posed a challenge to understanding what constitutes association and assembly in cyberspace. At first glance, FoAA online is predicated on access to the internet, impacting individuals from exercising their human rights online by excluding them from the internet (Voule 2019). Similarly, internet censorship and internet shutdowns also represent ways in which FoAA online can be restricted without recourse to AI (Voule 2019). Indeed, UN Human Rights Council (HRC) resolution A/HRC/RES/38/11 affirms the importance of basic internet access as an element of assembly and association, being concerned with: 'undue restrictions preventing Internet users from having access to or disseminating information at key political moments, with an impact on the ability to organize and conduct assemblies' (UN Human Rights Council 2018a, 2).

Beyond internet access, the UN HRC have supported and affirmed that FoAA also applies once an individual has access to the internet. In multiple UN HRC resolutions (resolutions 20/8, 26/13, 32/13, 38/7) dating to 2012, 'the promotion, protection, and enjoyment of human rights on the Internet' has been affirmed (UN Human Rights Council 2012; 2014; 2016; 2018b). This applies to FoAA online, which has been explicitly highlighted under A/HRC/RES/24/5 which reiterated:

the important role of new information and communications technologies in enabling and facilitating the enjoyment of the rights to freedom of peaceful assembly and of association, and the importance for all States to promote and facilitate access to the Internet and international cooperation aimed at the development of media and information and communications facilities in all countries (UN Human Rights Council 2013, 2).

Further, in 2018, UN HRC resolution A/HRC/RES/38/11 expressed 'that although an assembly has generally been understood as a physical gathering of people, human rights protections, including for the rights to freedom of peaceful assembly, of expression and of

association, may apply to analogous interactions taking place online' (UN Human Rights Council 2018a, 2). Finally, UN General Assembly resolution A/RES/73/173 in 2018 called upon 'States to ensure that the same rights that individuals have offline, including the rights to freedom of expression, of peaceful assembly and of association, are also fully protected online' and further, to grant 'respect to all individuals exercising their rights to freedom of peaceful assembly, of expression and of association, online and offline, in cases of threat, harassment, violence, discrimination, racism and other violations and abuses committed against them' (UN General Assembly 2018, 3).

FoAA online vs. offline

The importance of formal UN acknowledgment of FoAA online should not be understated. But what is FoAA online, and how does it differ from assembly and association offline? The United Nations resolutions do not emphasise specific platforms, but focus on the fact that states have obligations to protect these rights online regardless of platform (Voule 2019). From the perspective of the human rights mechanisms at the United Nations, assembly and association on the internet can be seen as any use of the internet to exercise FoAA. More specifically, it can be understood as individuals coming together and/or forming groups peacefully in a public or private online space for some purpose (Association for Progressive Communications 2019; Australian Government: Attorney-General's Department 2015). This can include, but is not limited to, using social media, organising a protest, DDoS, gathering signatures, fundraising, meeting on forums and more (Council of Europe 2015). The focus is on the ability of individuals to come together online – differing from what they say and how they say it, which more closely falls under freedom of expression.

FoAA online focuses on how the rights to assembly and association are used on the internet, whether it is entirely online, such as online petitions, or enabled by the internet, such as using social media to organise an offline protest. The use of the internet to organise protests and call for change has been well-known since the 2009 Iranian Green Movement popularised the phrase 'Twitter Revolution' and became 'the first major world event broadcast worldwide almost entirely via social media' (Keller 2010). More recently, a petition to repeal Brexit brought over 6 million people together to sign the largest petition ever delivered to parliament (Cheung 2019). Finally, The #MeToo movement saw millions of women assemble online to share their stories of sexual harassment, resulting in an ongoing collective conversation in many countries about the topic (Ghadery 2019). In all these cases, and many more, the internet was essential to these movements gathering significant attention and allowing millions of individuals to come together.

Beyond organising protests and assembling for change, FoAA online can restructure how we understand assembly and association: individuals can meet in person to organise and have others attend via Skype. Alternatively, a spontaneous protest movement, such as Iran's Green Movement, moves from offline to online and forms the basis for long-lasting association and assembly structured around political change (Niknejad 2013). Online petitions, WhatsApp groups, email, social media, 'meetup' apps and other similar applications and platforms are utilised to come together, share information, discuss, organise, advocate and other actions related to peaceful assembly and association entirely online, as a conduit to offline assembly and association, or in hybrid forms such as live-tweeting offline protests.

Further, in some political and social contexts, utilising the internet may be the only way assembly and association can be exercised. In dangerous contexts, or when assembly and association may endanger lives or be banned, the internet is often a vitally important way to exercise these human rights. On a more practical level, online FoAA enables individuals with physical disabilities, or those who cannot afford to travel to protests or meetings, to exercise assembly and association in ways which are supportive of their circumstances. Indeed, it is challenging to separate the use of the internet from various scales of FoAA and create a distinctly online and offline dichotomy. Nonetheless, challenges and threats to FoAA, unique to the internet, do exist.

Threats to FoAA online

In his recent 2019 report, UN Special Rapporteur on the rights to freedom of peaceful assembly and of association, Clément Nyaletsossi Voule, stated that he 'is concerned about the variety of measures and tactics that are used by States to control and impede access to and use of digital technology for the exercise of the rights to freedom of assembly and of association' (Voule 2019, 8). Not limiting his concerns to states, Voule (2019) argued that technology companies also share responsibility under the *Guiding Principles on Business and Human Rights* to ensure that they do not violate the human rights of their users. Voule identified several key areas which states must address: laws restricting or limiting online behaviour or content related to assembly and association, content blocking, state-sponsored cyberattacks, internet shutdowns or disruptions, internet taxes and surveillance. He also identified content display and moderation as an area where technology companies can work to ensure that the rights to association and assembly are upheld.

Of the threats Special Rapporteur Voule noted, AI is currently used in content display and moderation and surveillance (Kaye 2019; Feldstein 2019b). Although recent articles have discussed the potential role for AI in cyberattacks, these types of attacks have not yet manifested in a significant way (Lynch 2017). Thus, content display and moderation and surveillance represent the key vectors in which AI will impact FoAA in the future.

Content display

Content display occurs when algorithms determine what content appears in search results or on social media feeds. AI systems learn which content generates user interactions, and the system will respond to this data based on how it has been programmed – most usually it serves more of the same type of content. For instance, Instagram's AI system will take as input liking photographs of flowers and respond to this pattern by delivering more photographs of flowers as well as frequently liked photographs by other users who liked flowers (Tufekci 2019b). The data can come from a single internet browsing session, years of social media use, or from anonymous aggregate data based on geography and other demographics (Diresta 2018).

The trend in content display is towards greater personalisation and is customisation driven, in part, by AI (Tufekci 2019b). Given the vast amount of content on the internet, this is understandable and allows individuals and organisations to organise content. However, in the process of organising content, algorithms promote certain types of content and influence what content is displayed. As a result, the internet is invisibly

influenced by algorithmic logic. Indeed, UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, David Kaye states: 'Artificial intelligence-driven personalisation may also minimise exposure to diverse views, interfering with individual agency to seek and share ideas and opinions across ideological, political or societal divisions' (Kaye 2018, 7).

It can be added that AI-driven personalisation may minimise how and where individuals assemble online, what types of associations can be formed and the extent to which the rights to both assembly and association are respected. AI may deprioritise content with low engagement and prefer content with higher engagement, meaning that 'savvy actors' can manipulate AI rules to have disproportionate influence online (Kaye 2018; Tufekci 2019a). The recent disinformation campaigns in the United States are a case in point, with Facebook's algorithmic logic being manipulated by Russia's Internet Research Agency to promote disinformation. The end result was hundreds of thousands of individuals following Facebook pages focused on anti-immigrant rhetoric and hate speech (Frenkel and Benner 2018; Sanger 2019). Some of these algorithmically promoted pages held offline events attracting hundreds of protestors in Houston, Texas (Lucas 2017). With billions of people using social media, the ways in which AI shapes the information environment directly impacts how assembly and association manifest.

Content moderation

In the face of the overwhelming amounts of content, major online platforms are turning to AI to provide moderation (Gollatz, Beer, and Katzenbach 2018; Kaye 2019). Automated content moderation uses AI to remove content, respond to user reports, mediate comments and delete spam. It is also used to recommend content for further review by human moderators. Here, AI can influence assembly online and offline by eliminating events or conversations from social media. This can be more harmful than content display because it deprives individuals of the opportunity for assembly and association, rather than making it simply difficult to find. An internet search for a petition or protest which has been subject to content display will eventually turn up the event or group. A similar search for some event or petition which has been moderated will turn up no evidence of its existence.

Leading human rights and technology organisation, Access Now, defines content moderation as 'the practice through which an online service deals with user-generated speech' and decides what content should exist on a platform (Access Now 2019). This is different from content display, which is how AI determines what content appears on social media newsfeeds. The key difference is that content display prioritises certain content over others but doesn't remove content. Content moderation, on the other hand, determines what content *exists* on a platform.

Similar to training AI for content display, in content moderation, AI is trained by humans using samples of content to be moderated so it can learn what to remove or recommend for removal to humans (Kaye 2019). However, both the algorithms and training data used on social media platforms are unavailable for external examination and review (Kaye 2018). This opaqueness of content moderation, coupled with inadequate access to remedy, means that users are unable to know how or why content is being removed or hidden from them (Kaye 2018).

Despite assertions by major platforms that AI can resolve issues related to moderating content, such as hate speech, misogyny or radicalisation, increasingly research is demonstrating that AI-based content moderation is fallible. Two recent studies have shown that leading AI content moderation models for detecting hate speech are biased against African-Americans, being 1.5 times more likely to identify tweets written by African-Americans as being offensive, in addition to having other elements of significant racial bias against African-Americans and other vulnerable groups with different, non-Caucasian speech norms (Sap et al. 2019; Davidson, Bhattacharya, and Weber 2019).

When AI is not the primary decision maker for content moderation, it will recommend content for moderation to humans (Roberts 2019). These moderators exhibit all-too-human vulnerabilities and can lack local contextual understanding and language skills, and may be politically or socially biased (Cox and Koebler 2018). Further, human moderators often suffer from psychological burnout, post-traumatic stress disorder and other risks related to repeated exposure to disturbing content that can impair their ability to moderate content (Gillespie 2018; Newton 2019; Roberts 2019). The potentially serious implications of AI and human content moderation are evidenced by Facebook's role in the Rohingya disaster due to AI's inability to detect nuance in hate speech and a lack of locally qualified content moderators (Stecklow 2018).

According to Roberts (2019), humans will continue to play an important role in moderating content. She argues that the importance of AI is overstated, and that humans will never be fully taken out of their vital role in content moderation, at the very least for training the algorithms (Mack 2019). Duarte, Llanso, and Loup (2018) concur that in its present state of technological development, AI is not able to provide suitable content moderation alone because it is inaccurate and easily manipulated, among other issues. Thus, for the foreseeable future, moderation will be influenced by AI with final decisions largely in the hands of humans.

Nonetheless, even if final moderation decisions rest with a human, AI will recommend certain content for moderation over other content, as well as automatically regulating easily-identifiable content, which it is already doing (Wong 2019). Despite AI's documented shortcomings, platforms such as Google, Twitter and Facebook have made explicit commitments to increasing AI in content moderation, with Mark Zuckerberg stating that Facebook's AI will 'take action on the content automatically' (Ghaffary 2019; Zuckerberg 2018). Whether AI is solely responsible for moderation, or acts as a moderation-recommendation engine, the role of AI in content moderation is increasingly central to assembly and association online.

The arguments over the ineffectiveness of AI in content-moderation and the continued relevance of humans is important in understanding how these systems will impact FoAA online. However, while the usage of AI for content moderation will vary, and it may be ineffective, the significant media coverage of AI and its representation in popular culture, may disproportionately influence vulnerable populations who do not have access to the latest research demonstrating AI's fallibility or questioning its effectiveness (Bergstein 2017; Cave et al. 2018). That is, vulnerable and at-risk populations who use social media and the internet for assembly and association may not be willing to take any risk that they could be discovered, harassed or arrested because of AI. Indeed, the simple fact that content is moderated and removed may be enough to deter both online and offline assembly and association. In the case of content moderation, AI's

opaqueness, fallibility, lack of remedy and outsize media excitement, coupled with explicit commitments to increase AI moderation, mean there are multiple vectors through which AI can deprive users of their ability to sign petitions, organise protests and otherwise exercise their rights to assembly and association online (Ghaffary 2019).

Algorithmic censorship

Broadly speaking, the two categories of content display and moderation fall under the framework of ‘algorithmic censorship’. Sociologist Zeynep Tufekci coined this phrase to describe the phenomenon of algorithms determining what we can or cannot see, and the degree to which we can interact with that content (Tufekci 2015). It is not that the algorithms themselves are problematic, but rather that they are far from neutral (Tufekci 2015). Algorithms are influenced by the gender identity, socioeconomic class, culture, upbringing and other conscious and unconscious biases of their developers, through direct programming or by the data used to train the algorithms. These variables are a given in any sociotechnical project, but with the vast amount of information online and the role which algorithmic censorship plays in social media, these variables will play a decisive role in the future of assembly and association online.

As an example of how unconscious bias can manipulate content, Google’s image AI identified African-Americans as gorillas (Zhang 2015). With less than 2% of the technical workforce at Google being black (Nisen 2014), an unconscious programmatic bias by engineers could have allowed this classification. Beyond this, the question of training data and bias in that data would likewise influence these results. For example, Google’s Cloud Natural Language system considered the phrase ‘I’m a Jew’ to have a negative meaning, having used faulty training data (Thompson, Matsakis, and Koebler 2017).

Thus, the opaqueness of AI development has the potential to be problematic for human rights. These dangers are compounded when AI is deployed in support of anti-blasphemy laws, to combat hate speech, or to defend against terrorism. Algorithmic censorship has the potential to create a gaslighting effect, with users unsure about what they see and why they see it, while at the same time lacking platform transparency and remedy to hold platforms accountable (Kaye 2018). With many large tech platforms already erring on the side of caution, and the open internet already on the decline (Freedom House 2018), and even assuming the best of intentions on behalf of AI development, the potential exists for AI to limit human rights with opaque algorithmic censorship.

AI and surveillance

Artificial intelligence is a key part of surveillance in authoritarian, hybrid and democratic regimes. The use of AI in surveillance is often framed as a positive – for instance, as of 2019, the bulk of AI’s impact on surveillance has been in ‘smart cities’ initiatives, facial recognition combined with closed-circuit television, and for crime analysis in service of smart policing (Feldstein 2019b). Making cities more responsive and promoting public security are arguably positive goals, and AI is touted to achieve these healthier and safer cities, however this has the unintended effect of normalising AI-based surveillance. Additionally, AI is actively used to comb through online data to identify potential terrorist threats (Ganor 2019), anticipate epidemics based on social media posts (Umali 2019), gather open source

intelligence (Eldridge, Hobbs, and Moran 2018), and enhance medical research (Nazha 2019). While these are arguably socially beneficial, they also insert artificial intelligence directly into multiple surveillance regimes.

In fact, Freedom House reports that AI is being used by governments to pre-emptively identify threats, monitor dissent and silence dissidents (Freedom House 2019; Feldstein 2019a). Through social media surveillance, AI can be used to monitor sentiment in comments, map a user's network of friends, harass journalists (Reporters Without Borders 2018) and more (Freedom House 2019). Governments are partnering with startups, such as Clearview AI, to develop surveillance systems which utilise AI to match photographs of unknown people (such as a random snapshot on the street) with publicly scraped photographs from social media and other personal information online (Hill 2020). Already, Clearview AI's database of over 3 billion photos is available to the US Federal Bureau of Investigations and local law enforcement and has been used to 'search by face' to identify and arrest criminals (Hill 2020).

As these types of surveillance become ubiquitous, the online spaces where individuals assemble and associate can become places of vulnerability. State AI-facilitated surveillance of social media and Clearview AI's products are early examples of AI-enabled internet surveillance. For instance, with Clearview AI, law enforcement can leverage AI to automatically locate and cross-reference online social media profiles and other information for people at a protest – deterring both online and offline assembly and association (Hill 2020). As a result, individuals may become less likely to attend in-person and online gatherings. As AI development proliferates, and as AI use in social media becomes normalised through content display and moderation efforts, policy experts in academia, think tanks, government and elsewhere must develop proactive policy solutions.

Policy responses

It is important to note that there is considerable alarmism surrounding AI. The recent debunking of the myth that YouTube's algorithm encourages radicalisation is an example of how AI may not be having the detrimental effects on platforms which alarmists have believed (Ledwich and Zaitsev 2019). Nonetheless, AI will greatly impact online behaviour and engagement, either through its actual workings, alarmism and fear, or utopianism.

Academics and other experts have notably examined the impact of AI on privacy (Privacy International and Article 19 2018; Raso et al. 2018) and freedom of expression (Kaye 2018). However, the rights to assembly and association are vital elements of democratic societies and yet the literature on this right and AI's impact is lacking. To remedy this and contribute to a more well-rounded body of literature about AI and human rights, this paper will present some preliminary policy recommendations.

The threats to assembly and association from artificial intelligence are clear, but the remedy is less so. Fortunately, there is a critical upswelling of interest in the AI industry, social media companies and civil society to ensure that AI is developed with greater regard for its societal impacts, resulting in over 60 public, private or public-private AI ethics initiatives (Mittelstadt 2019). Recent examples include the *Toronto Declaration* for AI development using international human rights law as a framework, the *Santa Clara*

Principles on Transparency and Accountability in Content Moderation, the industry consortium The Partnership on AI, and the World Economic Forum's AI Council.

The pressure from civil society has manifested in various AI ethics boards to advise and guide companies in their AI development. However, a purely ethical approach can lack clarity and lead to ethically sound development which nonetheless violates human rights (Hidvegi and Leufer 2019). An ethical approach may, for instance, focus on culturally-specific ethical frames, creating standards of protection and care which vary by geography. It may be ethical to ensure that AI limits association online to minimise the potential for protest movements, but does that ethical frame respect human rights? Hidvegi and Leufer argues that:

While ethical reasoning and individual conscience certainly have a role to play, we believe that it is important to stress the role of human rights here. Beyond their embodiment in specific laws, human rights offer us a broad and well-defined set of principles to cover all instances in which our dignity and integrity are threatened (Hidvegi and Leufer 2019).

Any approach to protecting FoAA online should incorporate both advances in the field of AI ethics as well as well-established human rights law. Despite the fact that states are the duty bearers for human rights, much of assembly and association online takes place on the platforms of private technology companies and the development of AI is being driven by private companies (Kaye 2018). Thus, the efforts to protect assembly and association must be borne by both states and technology companies.

States must ensure that laws will not unduly impact FoAA online, except in clearly demarcated exceptions in line with existing human rights law (UN General Assembly 1966). This must include rethinking of FoAA as it applies to the internet, for example, understanding that legislation banning encryption or encouraging the use of 'real names' in social media could prohibitively impact FoAA. Further, states must consider legislation that would not harm, but proactively protect, users, such as protection from automated trolling for women (Van Der Wilk 2018).

Recognising the importance of technology companies, states should use their legislative ability and influence to encourage AI industry standards and development principles that will support competition in the development of AI, moving AI out of the hands of the 'Big 5': Facebook, Google, Microsoft, Amazon and Apple (Sawers 2019). Encouraging competition in AI development can support alternative norms and bring broader pressure to ensure human rights standards are incorporated in AI (Kaye 2018).

The private nature of AI development makes verification of human rights standards difficult, although some in the internet technical community have embraced incorporating human rights into technical protocols (Milan and Ten Oever 2017). States must encourage and support expert panels and code audits with the necessary protections for intellectual property. Various forms of code review can be developed which encourage human rights review while protecting corporate intellectual property (Kaye 2018).

Beyond states, corporations should follow the *UN Guiding Principles on Business and Human Rights* and ensure that their artificial intelligence code respects assembly and association. In order to achieve this, technology companies should conduct regular human rights impact assessments (HRIAs). HRIAs for AI and FoAA should include outside experts which can review AI for ways in which it may impact the ability of individuals to come together and organise online.

To adequately do so, HRIAs should examine how AI is developed, the training data used, and what the results of testing reveal about how it will impact diverse individuals and groups to come together. Some key questions to examine would be:

- What kind of data is used to train the AI and how was that data collected? What might the impact of using this data be on the ability of different types of individuals and groups to come together in the target population?
- What groups, perspectives, politics, languages, religion, levels of physical ability, etc. are represented in the training data, who was excluded, why were they excluded, and how does that compare to the target population where this AI will be deployed?
- Errors in code are inevitable. What types of errors have been detected in the AI during testing, and how did these errors impact the test population? How will these errors impact the live target population?
- What types of content is the focus of this AI? How was this content defined and delimited in theory, training and practice?

The questions aim to examine how AI training and testing will impact people and groups coming together, how the algorithm recommends content, and who will be impacted and how by those recommendations. For example, does the AI support assembly and association by disabled individuals, elderly, or other vulnerable, marginalised, or minority groups? What kind of assemblies and association does it encourage, intentionally, through programming, or unintentionally, with errors?

These recommendations are starting points for policy solutions. The recommendations offered should be considered within a broader HRIA by companies that ensure that other human rights are respected. Given the limitations of online platforms, however, it is understood that not all rights can be perfectly protected and guaranteed. To that end, Safiya Noble (2018) argues that discrimination, in some form, is part of modern technological development. Indeed, the nature of algorithmically managing social media feeds requires favouring some content or groups over others, thus some FoAA violations will be present irrespective of no matter the best intentions. To address this, HRIAs should examine and make clear how and why certain content is promoted or demoted and how users can appeal decisions and what that process entails. The lack of remedy has been specifically highlighted as a major shortcoming of extant AI implementation and yet is one of the most feasible ways to enhance human rights protections in AI (Kaye 2018).

In summary, this paper makes the following policy recommendations:

- Both state and corporate efforts to protect FoAA online should incorporate advances in the field of AI ethics and international human rights law.
- States should:
 - Understand how even well-intentioned laws can specifically and uniquely impact FoAA online and ensure that laws will not unduly harm FoAA online;
 - Use legislative powers to influence and encourage AI industry standards informed by human rights law;
 - Support diversity in the AI development ecosystem by encouraging competition that would help move AI development out of the hands of large tech companies; and

- Encourage the use of AI expert panels for human rights-focused AI code audits, with necessary protections for corporate intellectual property.
- Corporations should:
 - Follow the UN Guiding Principles on Business and Human Rights when developing and deploying AI; and
 - Conduct regular Human Rights Impact Assessments (HRIAs) for AI and FoAA. These HRIAs should address key questions on how AI is developed, the training data used, how AI recommends content, and how AI will impact FoAA online for diverse individuals and groups.

Conclusion

The internet is a collective space where human rights are exercised and restricted for billions of people every day. Vulnerable groups and individuals often use the internet as the sole space in which they can exercise their human rights, including the right to assembly and association. Increasing digital authoritarianism has made it difficult for many of these individuals and groups to exercise their basic human rights online (Frenkel and Benner 2018), impacting their sense of self and human agency (Ashraf 2015).

The emergence of artificial intelligence, while offering numerous benefits, symbolises a critical juncture in the evolution of the internet as collective space. With the potential for human rights backsliding due to algorithmic censorship, users should be aware of AI's impact on assembly and association, alongside other human rights. Whether intentional or unintentional, AI is both enhancing and limiting this space and has the potential to actively erode human rights by de-ranking or removing content. While the volume of content on the internet requires some type of sorting by AI, how it does this, who it impacts, and what rights and responsibilities are coded into AI remain a mystery and must be more transparent.

Despite well-meaning AI ethics boards, AI development must incorporate, or at a minimum, seriously consider, the human rights implications of AI development and implementation. The internet is no longer simply a space to receive and impart information, but a place where many find the solace in coming together with our fellow humans. It is vital that civil society, academia, government and industry likewise come together to examine the role human rights principles can play in the future of AI.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributor

Cameran Ashraf is assistant professor of new media and global communications at the School of Public Policy, Central European University and co-founder of AccessNow, now one of the largest international human rights organisations dedicated to defending human rights in the digital age. Based on his work, the European Parliament selected AccessNow as finalist for the 2010 Sakharov Prize for Freedom of Thought, the European Union's highest human rights honour.

His international human rights and technology work has included defending critical websites from state-sponsored attacks, providing personal communications security for vulnerable activists and journalists, distributing anti-censorship tools used by thousands of people daily, and facilitating more than 3 million video downloads from inside authoritarian regimes.

He has been invited to speak at Harvard University, MIT, UC Berkeley, the OSCE and has appeared in the New York Times, Wired Magazine, and Bloomberg Businessweek.

ORCID

Cameran Ashraf  <http://orcid.org/0000-0003-3275-3624>

References

- Access Now. 2019. "Protecting Free Expression in the Era of Online Content Moderation." *Access Now*. <https://www.accessnow.org/cms/assets/uploads/2019/05/AccessNow-Preliminary-Recommendations-On-Content-Moderation-and-Facebooks-Planned-Oversight-Board.pdf>.
- Ashraf, Cameran. 2015. "Who Do We Lose With Internet Censorship and Control? · Global Voices." *Global Voices* (blog), February 4. <https://globalvoices.org/2015/02/04/who-do-we-lose-with-internet-censorship-and-control/>.
- Association for Progressive Communications. 2019. "The Rights to Freedom of Peaceful Assembly and of Association in the Digital Age: APC Submission to the United Nations Special Rapporteur on the Rights to Freedom of Peaceful Assembly and of Association." Association for Progressive Communications.
- Australian Government: Attorney-General's Department. 2015. "Right to Freedom of Assembly and Association." Right to Freedom of Assembly and Association. October 22. <https://www.ag.gov.au/RightsAndProtections/HumanRights/Human-rights-scrutiny/PublicSectorGuidanceSheets/Pages/Righttofreedomofassemblyandassociation.aspx>.
- Belyaeva, Nina, Thomas Bull, David Goldberger, Michael Hamilton, Neil Jarman, Muatar Khaidarova, Serghei Ostaf, Vardan Poghosyan, Alexander Vashkevich, and Yevgeniy Zhovtis. 2010. "Guidelines on Freedom of Peaceful Assembly." OSCE Office for Democratic Institutions and Human Rights.
- Bergstein, Brian. 2017. "Fiction That Gets AI Right." *MIT Technology Review*, October 24. <https://www.technologyreview.com/s/609131/fiction-that-gets-ai-right/>.
- Cave, Stephen, Claire Craig, Kanta Dihal, Sarah Dillon, Jessica Montgomery, Bethany Singler, and Lindsay Taylor. 2018. "Portrayals and Perceptions of AI and Why They Matter." The Royal Society.
- Cheung, Helier. 2019. "Brexit Debate: Do Petitions Ever Work?" *BBC News*, March 26. <https://www.bbc.com/news/world-47693506>.
- Council of Europe. 2015. "Report by the Committee of Expertson Cross-Border Flow of Internet Traffic and Internet Freedom on Freedom of Assembly and Association on the Internet." MSI-INT (2014)08 rev6 FINAL. Council of Europe.
- Cox, Jason, and Joseph Koebler. 2018. "Here's How Facebook Is Trying to Moderate Its Two Billion Users." *Vice* (blog), August 23. https://www.vice.com/en_us/article/xwk9zd/how-facebook-content-moderation-works.
- Davidson, Thomas, Debasmitta Bhattacharya, and Ingmar Weber. 2019. "Racial Bias in Hate Speech and Abusive Language Detection Datasets." In *Proceedings of the Third Workshop on Abusive Language Online*, 25–35. Florence: Association for Computational Linguistics. <https://www.aclweb.org/anthology/W19-3504>. doi:10.18653/v1/W19-3504.
- Diresta, Renee. 2018. "The Web's Recommendation Engines Are Broken. Can We Fix Them?" *Wired*, April 11. <https://www.wired.com/story/creating-ethical-recommendation-engines/>.
- Duarte, Natasha, Emma Llanoso, and Anna Loup. 2018. "Mixed Messages? The Limits of Automated Social Media Content Analysis." In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, edited by Sorelle A. Friedler and Christo Wilson, 81:106–106. Proceedings of Machine Learning Research. New York, NY, USA: PMLR. <http://proceedings.mlr.press/v81/duarte18a.html>.

- Eldridge, Christopher, Christopher Hobbs, and Matthew Moran. 2018. "Fusing Algorithms and Analysts: Open-Source Intelligence in the Age of 'Big Data.'" *Intelligence and National Security* 33 (3): 391–406. doi:10.1080/02684527.2017.1406677.
- Feldstein, Steven. 2019a. "The Road to Digital Unfreedom: How Artificial Intelligence Is Reshaping Repression." *Journal of Democracy* 30 (1): 40–52.
- Feldstein, Steven. 2019b. "The Global Expansion of AI Surveillance." Carnegie Endowment for International Peace.
- Freedom House. 2018. "Freedom on the Net 2018: The Rise of Digital Authoritarianism." *Freedom House*, October 17. https://freedomhouse.org/sites/default/files/FOTN_2018_Final.pdf.
- Freedom House. 2019. "Freedom on the Net 2019: The Crisis of Social Media." *Freedom House*, November 4. <https://www.freedomonthenet.org/report/freedom-on-the-net/2019/the-crisis-of-social-media>.
- Frenkel, Sheera, and Katie Benner. 2018. "To Stir Discord in 2016, Russians Turned Most Often to Facebook." *The New York Times*, July 16. <https://www.nytimes.com/2018/02/17/technology/indictment-russian-tech-facebook.html>.
- Ganor, Boaz. 2019. "Artificial or Human: A New Era of Counterterrorism Intelligence?" *Studies in Conflict & Terrorism*, February, 1–20. doi:10.1080/1057610X.2019.1568815.
- Ghadery, Farnush. 2019. "#Metoo – Has the 'Sisterhood' Finally Become Global or Just Another Product of Neoliberal Feminism?" *Transnational Legal Theory* 10 (2): 252–274.
- Ghaffary, Shirin. 2019. "The Algorithms That Detect Hate Speech Online Are Biased against Black People - Vox." *Vox: Recode*. August 15. <https://www.vox.com/recode/2019/8/15/20806384/social-media-hate-speech-bias-black-african-american-facebook-twitter>.
- Gillespie, Tarleton. 2018. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. New Haven: Yale University Press.
- Gollatz, Kirsten, Felix Beer, and Christian Katzenbach. 2018. "The Turn to Artificial Intelligence in Governing Communication Online." Alexander von Humboldt Institute for Internet and Society & Access Now.
- Hidvegi, Fanny, and Daniel Leufer. 2019. "Laying down the Law on AI: Ethics Done, Now the EU Must Focus on Human Rights." *Access Now* (blog), April 8. <https://www.accessnow.org/laying-down-the-law-on-ai-ethics-done-now-the-eu-must-focus-on-human-rights/>.
- Hill, Kashmir. 2020. "The Secretive Company That Might End Privacy as We Know It." *The New York Times*, January 18. <https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html>.
- Kaye, David. 2018. "Promotion and Protection of the Right to Freedom of Opinion and Expression." A/73/348. United Nations.
- Kaye, David. 2019. *Speech Police: The Global Struggle to Govern the Internet*. New York: Columbia Global Reports.
- Keller, Jared. 2010. "Evaluating Iran's Twitter Revolution." *The Atlantic*. June 18, 2010. <https://www.theatlantic.com/technology/archive/2010/06/evaluating-irans-twitter-revolution/58337/>.
- Kiai, Maina. 2012. "Report of the Special Rapporteur on the Rights to Freedom of Peaceful Assembly and of Association, Maina Kiai." A/HRC/20/27. United Nations. <https://www.bbc.com/news/world-47693506>.
- Ledwich, Mark, and Anna Zaitsev. 2019. "Algorithmic Extremism: Examining YouTube's Rabbit Hole of Radicalization." *ArXiv:1912.11211 [Cs]*, December. <http://arxiv.org/abs/1912.11211>.
- Lucas, Ryan. 2017. "How Russia Used Facebook To Organize 2 Sets Of Protesters." *National Public Radio*. November 1, 2017. <https://www.npr.org/2017/11/01/561427876/how-russia-used-facebook-to-organize-two-sets-of-protesters>.
- Lynch, Mike. 2017. "AI Cyberattacks Will Be Almost Impossible for Humans to Stop." *Wired UK*, December 28. <https://www.wired.co.uk/article/ai-cyberattack-mike-lynch>.
- Mack, Zachary. 2019. "Why AI Can't Fix Content Moderation." *The Verge*, July 2. <https://www.theverge.com/2019/7/2/20679102/content-moderation-ai-social-media-behind-the-screen-sarah-t-roberts-vergecast>.

- Milan, Stefania, and Niels Ten Oever. 2017. "Coding and Encoding Rights in Internet Infrastructure." *Internet Policy Review* 6 (1), <https://policyreview.info/articles/analysis/coding-and-encoding-rights-internet-infrastructure>. doi:10.14763/2017.1.442.
- Mittelstadt, Brent. 2019. "AI Ethics – Too Principled to Fail?" SSRN Scholarly Paper ID 3391293. Rochester, NY: Social Science Research Network. <https://papers.ssrn.com/abstract=3391293>.
- Nazha, Aziz. 2019. "Does AI Have a Place in Medicine?" *Scientific American*, November 11. <https://blogs.scientificamerican.com/observations/does-ai-have-a-place-in-medicine/>.
- Newton, Casey. 2019. "Google and YouTube Moderators Speak out on the Work That Gave Them PTSD." *The Verge*, December 16. <https://www.theverge.com/2019/12/16/21021005/google-youtube-moderators-ptsd-accenture-violent-disturbing-content-interviews-video>.
- Niknejad, Kelly Golnouth. 2013. "It Ain't Easy Being Green." *Foreign Policy* (blog), June 3. <https://foreignpolicy.com/2013/06/03/it-aint-easy-being-green/>.
- Nisen, Max. 2014. "Only 2% of Google's American Workforce Is Black." *The Atlantic*, May 29. <https://www.theatlantic.com/business/archive/2014/05/only-2-percent-of-googles-american-workforce-is-black/371805/>.
- Noble, Safiya. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. 1st ed. New York: NYU Press.
- Privacy International, and Article 19. 2018. "Privacy and Freedom of Expression In the Age of Artificial Intelligence." Privacy International & Article 19.
- Raso, Filippo A., Hannah Hilligoss, Vivek Krishnamurthy, Christopher Bavitz, and Levin Kim. 2018. "Artificial Intelligence & Human Rights: Opportunities & Risks." Berkman Klein Center for Internet & Society at Harvard University.
- Reporters Without Borders. 2018. "Online Harassment of Journalists: Attack of the Trolls." Reporters Without Borders.
- Roberts, Sarah T. 2019. *Behind the Screen: Content Moderation in the Shadows of Social Media*. New Haven: Yale University Press.
- Royal Society. 2017. "Machine Learning: The Power and Promise of Computers That Learn by Example: An Introduction." Royal Society.
- Sanger, David E. 2019. *The Perfect Weapon: War, Sabotage, and Fear in the Cyber Age*. Reprint ed. New York: Broadway Books.
- Sap, Maarten, Dallas Card, Saadia Gabriel, Yejin Choi, and Noah A. Smith. 2019. "The Risk of Racial Bias in Hate Speech Detection." In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 1668–78. Florence, Italy: Association for Computational Linguistics. <https://www.aclweb.org/anthology/P19-1163>. doi:10.18653/v1/P19-1163.
- Sawers, Paul. 2019. "How the 'Big 5' Bolstered Their AI Through Acquisitions in 2019." *Venture Beat* (blog), December 23. <https://venturebeat.com/2019/12/23/how-the-big-5-bolstered-their-ai-through-acquisitions-in-2019/>.
- Smith, Rhona, and Christien van den Anker. 2005. *The Essentials of Human Rights*. 1st ed. London: Routledge.
- Stecklow, Steve. 2018. "Why Facebook Is Losing the War on Hate Speech in Myanmar." *Reuters*, August 15. <https://www.reuters.com/investigates/special-report/myanmar-facebook-hate/>.
- Thompson, Andrew, Louise Matsakis, and Jason Koebler. 2017. "Google's Sentiment Analyzer Thinks Being Gay Is Bad." *Motherboard* (blog), October 25. https://motherboard.vice.com/en_us/article/j5jmj8/google-artificial-intelligence-bias.
- Tufekci, Zeynep. 2015. "Algorithmic Harms Beyond Facebook and Google: Emergent Challenges of Computational Agency." *Colorado Technology Law Journal* 13: 203.
- Tufekci, Zeynep. 2019a. "YouTube's Recommendation Algorithm Has a Dark Side." *Scientific American*, April 1. doi:10.1038/scientificamerican0419-77.
- Tufekci, Zeynep. 2019b. "How Recommendation Algorithms Run the World." *Wired*, April 22. <https://www.wired.com/story/how-recommendation-algorithms-run-the-world/>.
- Umali, Teresa. 2019. "Utilising AI and Social Media for Early Disease Outbreak Detection." *OpenGov Asia* (blog), November 28. <https://www.opengovasia.com/utilising-ai-and-social-media-for-early-disease-outbreak-detection/>.

- UN General Assembly. 1966. "International Covenant on Civil and Political Rights", 16 December 1966, United Nations, Treaty Series, vol. 999, p. 171.
- UN General Assembly. 2018. "Promotion and Protection of Human Rights and Fundamental Freedoms, Including the Rights to Peaceful Assembly and Freedom of Association." A/RES/73/173.
- UN Human Rights Council. 2012. "The Promotion, Protection and Enjoyment of Human Rights on the Internet." A/HRC/RES/20/8.
- UN Human Rights Council. 2013. "The Rights to Freedom of Peaceful Assembly and of Association." A/HRC/RES/24/5.
- UN Human Rights Council. 2014. "The Promotion, Protection and Enjoyment of Human Rights on the Internet." A/HRC/RES/26/13.
- UN Human Rights Council. 2016. "Resolution Adopted by the Human Rights Council on 1 July 2016 - 32/13. The Promotion, Protection and Enjoyment of Human Rights on the Internet." A/HRC/RES/32/13.
- UN Human Rights Council. 2018a. "The Promotion and Protection of Human Rights in the Context of Peaceful Protests." A/HRC/RES/38/11.
- UN Human Rights Council. 2018b. "The Promotion, Protection and Enjoyment of Human Rights on the Internet." A/HRC/RES/38/7.
- United Nations. 1948. "Universal Declaration of Human Rights".
- Van Der Wilk, Adriane. 2018. "Cyber Violence and Hate Speech Online against Women." PE 604.979. European Parliament.
- Vincent, James. 2018. "This Is When AI's Top Researchers Think Artificial General Intelligence Will Be Achieved." *The Verge*, November 27. <https://www.theverge.com/2018/11/27/18114362/ai-artificial-general-intelligence-when-achieved-martin-ford-book>.
- Vincent, James. 2019. "The Problem with AI Ethics." *The Verge*, April 3. <https://www.theverge.com/2019/4/3/18293410/ai-artificial-intelligence-ethics-boards-charters-problem-big-tech>.
- VivaTech. 2018. *Mark Zuckerberg, Chairman and Chief Executive Officer of Facebook Interview - VivaTech Conference*. VivaTech Conference. <https://www.youtube.com/watch?v=Gp0BnrHn6cA>.
- Voule, Clément Nyaletsossi. 2018. "Report of the Special Rapporteur on the Rights to Freedom of Peaceful Assembly and of Association." A/73/279. United Nations General Assembly.
- Voule, Clément Nyaletsossi. 2019. "Rights to Freedom of Peaceful Assembly and of Association: Report of the Special Rapporteur on the Rights to Freedom of Peaceful Assembly and of Association." A/HRC/41/41. United Nations General Assembly Human Rights Council.
- Webster, Graham, Rogier Creemers, Paul Triolo, and Elsa Kania. 2017. "China's Plan to 'Lead'in AI: Purpose, Prospects, and Problems." New America Foundation.
- White House. 2016. "Preparing for the Future of Artificial Intelligence." Executive Office of the President, National Science and Technology Council, Committee on Technology.
- Wong, Queenie. 2019. "Facebook's AI Helps Block or Remove 1 Million Accounts Each Day." *CNET*. <https://www.cnet.com/news/facebook-ai-helps-social-network-block-or-remove-one-million-accounts-a-day/>.
- Zhang, Maggie. 2015. "Google Photos Tags Two African-Americans As Gorillas Through Facial Recognition Software." *Forbes*, July 1. <https://www.forbes.com/sites/mzhang/2015/07/01/google-photos-tags-two-african-americans-as-gorillas-through-facial-recognition-software/>.
- Zuckerberg, Mark. 2018. "A Blueprint for Content Governance and Enforcement | Facebook." *Mark Zuckerberg's Facebook Blog* (blog), November 15. <https://www.facebook.com/notes/mark-zuckerberg/a-blueprint-for-content-governance-and-enforcement/10156443129621634/>.