
A COVFefe INQUIRY ON SOCIAL MEDIA DEPLATFORMING

Stefano Formicola

Department of Computer Science

Politecnico di Milano

Milano, Italy

stefano.formicola@mail.polimi.it

ABSTRACT

Covfefe refers to a viral tweet by Donald Trump when he was President of the United States, presumably being a typo, that instantly became one of his most popular tweets among others representing his controversial use of social media for official statements, that led companies owning these platforms to deplatform him after the United States Capitol attack in January 2021.

This work aims to analyze the event of Trump's deplatforming from an ethical perspective, arguing on the right to the freedom of expression and when it is acceptable to limit it, also discussing on the role that social media platforms have in digital modern democracies, its risks and why it needs to be regulated in name of the interests of all, starting from the invisibility factor framework in computer ethics. An introduction to philosophical and ethical bases in freedom of speech and content regulation is used to develop our arguments on the deplatforming of President Trump and the need for regulations of social media platforms to preserve a democratic ethos.

Keywords Social networks · Ethics · Censorship · Democracy

1 Introduction

On January 6, 2021, a session of the United States Congress working on the counting of votes of US 2020 Elections was interrupted by an assault and take-over of the US Capitol. These rioters were President Donald Trump supporters who were contesting the irregularity of the elections in favor of boycotting Trump's victory, hence supporting this theory and asking the Congress to refuse the proclamation of Joe Biden as winner of the elections to become President of the United States. These people were incited by Trump himself to march towards the Capitol to fight with strength to take back their victory; later that day, when riots already started, Trump tweeted other messages thanking the rioters and blaming Vice President Pence accusing him of betrayal. This event led to five deaths and at least 400 people being accused of crimes. After the certification of electoral votes was completed, in the morning of January 7, Trump tweeted messages of violence and didn't condemn the riot, thus leading Twitter, Facebook, Reddit, YouTube and other platforms¹ to deplatform President Trump silencing his accounts, in some cases with a permanent ban.

This paper aims to answer the question *Should political deplatforming be allowed in a democratic society?* taking as case study President Trump's deplatforming and focusing on the ethical implications and liability of this action by private corporations. In the following section we first provide a theoretical background to the freedom of speech, later introducing past and modern critiques to social media and the moderation of contents on internet; secondly we discuss and analyze the deplatforming event, giving several ethical explanations on when and why it is dangerous for modern democratic societies and how it can be regulated in order to avoid biased conflicts of interests.

¹<https://www.axios.com/platforms-social-media-ban-restrict-trump-d9e44f3c-8366-4ba9-a8a1-7f3114f920f1.html>

2 Background

2.1 Freedom of Speech and its Limitations

Freedom of speech, along with its limitation, is a classical philosophical inquiry analyzed and discussed in many contexts for its democratic yet controversial nature. As the Article 19 of the *Universal Declaration of Human Rights* states:

Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers.

Free speech allows us to take part in public or private discussions, to pronounce opinions and express beliefs, everything exercising the same democracy that leads to this freedom of speech; its controversial drawback, though, is that it needs to be limited. Indeed, as Stanley Fish wrote, “*free speech in short, is not an independent value but a political prize*” (1994, 102) and this prize is a working democracy that has to be defended against speeches posing a threat to it.

John Stuart Mill was one of the most influential philosophers in the field of liberalism, contributing and working on the argument of individual liberty. In one of his most famous works, *On Liberty* (1859), Mill wrote that any doctrine, no matter how immoral it can be, should have the liberty to be professed by everyone, because the act by the entire mankind of silencing an individual is no more justified than, on the contrary, that single individual silencing the whole mankind for contrasting opinions. This freedom of speech is not only necessary for dignity of persons, but it also results in a diversity of arguments thus human faculties such as perception, judgment and moral preference are exercised in making choices, in order to be more aware of the multitude of options available.

Since Mill's statements were very powerful and strong he understood that some rules should be put in place to regulate people in a society, and these limitations are regulated by the *harm principle*: the only motivation for limiting a member of a democratic community against their will is to block the practicability of harming other members. It is worth noting – in order to use that principle with moderation – that harm is to be intended as a direct invasion or violation of the rights of a person and not an illegitimate harm as it can be a claim expressing a point of view.

A further explanation of how language can harm is using the speech act theory [Austin, 1975] in which speech acts are distinguished in *locutionary* acts, as the expression of a sentence, *illocutionary* acts, as expressing beliefs, and *perlocutionary* acts, that have effects on listeners. The harm principle allows freedom of speech to any kind of acts, but some limitations on perlocutionary acts are necessary when these can cause harm to others.

2.2 Social Networks and Ethics

Social Networks are one of the greatest revolutions in human history, because they reshaped the way people interact with society and, bidirectionally, how society evolves with political, individual and institutional activities offering a space for people to gather, express their opinions, collaborate, communicate with big audiences and build audiences, spending even more time. These spaces are where political discussions and diplomatic affairs are done, where businesses communicate with their customers, educators share knowledge and much more; for these reasons we cannot think of social networks as separated spaces from the physical one, because the physical space is influenced by the digital one. They are part of – indeed – an *infosphere* [Floridi, 2014], i.e. a blend of online and offline environments, and therefore they deserved attention from the community of philosophers of information technology and ethicists of computer science.

The first philosophers to question on the ethical implication of the social proposes of internet – and, indirectly, social networks – were Albert Borgmann and Hubert Dreyfus at the beginning of the century, when social networks didn't spread yet but chat rooms, newsgroups and online gaming were considered social media. Borgmann's critique on internet as social media are strongly influenced by Martin Heidegger's view of technology (*technological determinism*) as a mean that “tends to constrain and impoverish the human experience of reality in specific ways”, leading the German-American philosopher to claim that online social networks' glamorous virtuality is a “virtual fog” that obscures real human bonds: this is the concept of *hyperreality* such as an environment specifically and conveniently designed to offer a stylized version of ourselves, clashing with the disconnected organic reality when coming back to it. Even though this critique may have seemed ahead of time in 1992, nowadays it applies little in a society in which the line between online and offline is so thin that philosophers felt the need to define the term *onlife* [Floridi, 2015], describing the increasing pervasiveness of information and communication technologies in our life; what instead is contemporaneous is Borgmann's point of view is the possibility to cut off people from the online social environment whether their presence is perceived as burdensome, simply by executing screening devices that today we would define as *user ban* or *user block* on a platform.

Hubert Dreyfus, instead, embraces Kierkegaard's thought when developing his reflections on online society: according to the philosopher the online space is not the pleasant *hyperreality* that clashes with bodily reality, but a mere and deficient substitute of real life in which, though, the absence of embodied presence leads to immoderate behaviors by the users and therefore this constitutes a menace. According to Dreyfus online social spaces lack the exposure to risk our real identities and, therefore, "the Net frees people to develop new and exciting selves", playing with identities, commitment and meaning without risking the otherwise present consequences of acting in the same way in the physical environment. One may argue that modern social networks are based on the concept of giving an online identity to real-world entities (i.e. people, institutions, etc.) but still Dreyfus' argument applies to present issues such as hate speech, spreading of fake news, online frauds and unlawful acts, usually performed through an identity play or, otherwise, a less perceived risk than in a physical context.

In the last decade modern critiques and ethical concerns moved to the moral duty of social network providers and developers [Manders-Huits, 2010], whose product is involved in the autonomous construction of users' moral, political and practical identity therefore one of the common questions is "*can we trust these companies' recommendation algorithms?*". Indeed, the potential is that people could unconsciously be shielded with informational silos, namely *echo chambers* and *filter bubbles*, by social media in favor of a user experience that exposes users to a partial view of the reality – the one they most sympathize with – regardless of the integrity of the public sphere. The concern, then, is that this could promote extremism, ill-founded viewpoints and insurrections leading to damage democracy and freedom.

2.3 Internet Content Regulation

With content regulation on internet it is meant the censorship – by a content provider or an Internet Service Provider – of a certain content in order to stop its availability. It has been subject of study in philosophy of information and it is still a current issue given not only the different kind of media internet is, compared to traditional ones (e.g. television, books), but also whether the content on the internet *should* be regulated and, if so, by whom. One of the first scholars to question on this matter was John Weckert who not only criticized Mill's harm principle, but also pointed out that internet content regulation already exists (things that are illegal on other media are illegal on internet as well) and this regulation of content, to be effective, must be global [Weckert, 2000].

A modern scrutiny is however necessary when contents are available on content sharing platforms as social networks, especially with phenomena as the spreading of fake news or messages of hate. These interventions of content moderation show a biased position by social media companies because, otherwise, there couldn't be the resolution of disagreements given the prioritization of one set of values before others. Companies must, therefore, find a balance between the opportunity of free speech and the removal of harm contents, for instance by transparently clarifying on what bases their decisions are made and which problems they want to address [Stewart, 2021].

Deepening into the always more implication of social media into our lives (that are a core part of the *infosphere* and the *onlife* experience), a critical issue regards the sovereignty on these contents when social media are used by governments and political actors to communicate with the mass. Do these private companies have the right to manipulate the democratic debate on their platforms, based on economic interests, particularly when political leaders take part in it? In the next section we will analyze the issue and try to provide an answer to this ethical question.

3 Should Political Deplatforming Be Allowed in a Democratic Society?

If we look up for the definition of the term deplatforming we will find the following: *attempt to boycott a group or individual through removing the platforms (such as speaking venues or websites) used to share information or ideas*². Straightforwardly we can agree that deplatforming is a form of censorship, a limitation to the freedom of speech enunciated by the human right previously introduced and characterizing modern democratic societies. It is the case of what happened to President Trump, who has been deplatformed from all the main social network platforms receiving, in some cases, a permanent ban³, despite this has been celebrated as the victory of liberal democracy. There are however some ethical and moral concerns on this episode that made us reflect on the future implications and divergences from democratic values and politics; we will give an explanation of the moral bases that made this happen and provide an ethical point of view of why this action by a private company reveals a threat that ought to be regulated.

One may argue that political deplatforming is nothing different from the content moderation of a user's illocutionary act of expressing, for example, that a certain ethnic group should have more privileges than other ethnic groups. This is, in fact, the free expression of an opinion and – as far as unacceptable – we tolerate and welcome its censorship despite being a limitation to the freedom of speech; it has to be noticed that this is only a problem of content moderation and

²<https://en.wikipedia.org/wiki/Deplatforming>

³<https://edition.cnn.com/2021/01/08/tech/trump-twitter-ban>

has nothing to do with deplatforming. Private companies owning social network platforms have the right to indicate which kind of contents constitute a problem and should therefore be flagged for removal; from one side this “liberal democratic” bias allows a set of norms, a code of conduct, that users have to respect for a civic and beneficial stay on these platforms, on the other side there is a commercial interest in order to maximize users’ time increasing their activity, on which these platforms’ business models are based.

The deplatforming of Trump, instead, cannot be reduced in a matter of violation of terms of service: as we are moving towards an even more pervasiveness of social media in our life and society, these platforms have become like public plazas, where politics is done and communicated, sometimes replacing traditional media almost always anticipating them. This evolution gave these corporations a powerful decision-making role: on their platforms it is exercised the democratic development and transmission of the public discourse by governors, governments and institutions, and the fact that they can play such an unbalanced role in choosing whom and when can speak to a nation is an ethical issue.

We decided to adopt Moor’s invisibility factors [Moor, 1985] as a framework to provide three reasons, totally invisible to citizens of a democratic society, to explain why political deplatforming by a private company is ethically dangerous:

- *invisible abuse* – perhaps the easiest motivation, because each decision about the kind of contents available on these platforms may be moved by economic interests or other kinds of political motivations;
- *invisible programming values* – because in the development of a form of censorship some ethical questions may arise, as *what is right* or *what is wrong* for a ‘good’ survival of the democracy? Similarly to the programmer who makes some judgments in Moor’s example, the individual or group of people responsible for the deplatforming must give answers and take decisions on ethical concerns, replacing the judicial authority of a state;
- *invisible complex calculations* – because, in the end, when ethical dilemmas arise we are not aware of what assumptions or decisions led these companies to choose what is good and what is bad, and we need to worry about this potentially biased role.

At this point we may end up thinking that deplatforming Donald Trump was wrong, that his censorship must be condemned, but the harm principle comes at our help: limiting Trump’s free speech was right, because there was harm, violence and some people died while assaulting Capitol Hill. His tweets and videos were first an incitement, then a message of encouragement to violence and later a denial of his defeat over the elections; all his messages were of the kind of perlocutionary acts, inciting people to march towards their victory, against the right to democracy of all the other citizens but, more important, against the same democracy. At some point for these companies it became necessary to stop the President’s freedom of expression, whether for commercial reasons or for not being comfortable with their platforms being used to express certain intolerant ideas they don’t agree with; to some extent this reflects the paradox of tolerance described in *The Open Society and Its Enemies* [Popper, 1945]. In his work, philosopher Karl Popper argued that if society do not impose limits to tolerance, its ability to be tolerant is eventually destroyed by the intolerant: to rephrase it, maintaining a tolerant society means not tolerating intolerance.

“Unlimited tolerance must lead to the disappearance of tolerance. If we extend unlimited tolerance even to those who are intolerant, if we are not prepared to defend a tolerant society against the onslaught of the intolerant, then the tolerant will be destroyed, and tolerance with them. [...] We should therefore claim, in the name of tolerance, the right not to tolerate the intolerant.

Can we trust that next time these corporations will act according to the paradox of tolerance and harm principle, protecting society and its democratic values? Basing our reflections on the invisibility factors presented above and given the fact that they created a precedent of manipulation of the democratic speech, we claim that the digital society of the future needs more than deplatforming. An ethical set of laws is needed to regulate the role that these private companies have in public speech and in the *infosphere*, to ensure that next time they will act in the interest of democracy without any possible conflict of interests. As proposed by Floridi, we support the view that the infosphere – and therefore their main actors as social network companies – should be governed and regulated according to international laws like an international territory that belongs to everyone because “It is only there where those who control the controllers are the controlled themselves that one can talk of real democracy, and it is only in a real democracy that a limit to freedom of speech is not censorship but tolerant respect for civil communication, one that hurts nobody and is good for everybody” [Floridi, 2021]. For example, a human rights framework that ensures the transparency of decisions is the Manila Principles on Intermediary Liability [Electronic Frontier Foundation, 2015], based on international human rights and legal frameworks, protecting the freedom of expression and creating a transparent environment on how, when and why rules are applied.

In conclusion, we believe our thesis raises an important topic of concern for an unbiased evolution of democracy and society, and should therefore be taken in consideration for further analyses to avoid an uncontrollable slippery slope

effect of manipulation of public speech. We hope, instead, that governments will agree on a new set of international laws to protect and defend Institutions' online presence, in the name of the collectivity and democratic values, such as the same freedom of speech for everyone with which we started our discussion.

4 Conclusions

Social media undoubtedly played an important role in the assault and take-over of the US Capitol on January 6 for their decision to deplatform President Donald Trump. Since then a critical question is when it is necessary to deplatform a head of state and if this decision power belongs to these platform. In this paper we discussed on why political deplatforming should not be allowed in a democratic society relying solely to social network companies, we used the invisibility factor framework of computer ethics to answer this question and provided a theoretical explanation of why it happened, why it was right and why it constitutes a threat to modern democracies. In the end, we supported and introduced two complementary views on how to regulate this phenomenon in order to act, the next time, for the interest of all.

Acknowledgments

Part of the Background section of this paper, more specifically Sections 2.1 and 2.2 on Freedom of Speech [van Mill, 2021] and Social Networks and Ethics [Vallor, 2021], have been inspired from the extensive literature review presented in Stanford Encyclopedia of Philosophy⁴ and here cited to avoid repeated citations.

References

- [Austin, 1975] Austin, J. L. (1975). *How to do things with words*. Oxford university press.
- [Electronic Frontier Foundation, 2015] Electronic Frontier Foundation (2015). Manila principles on intermediary liability.
- [Fish, 1994] Fish, S. (1994). *There's no such thing as free speech: And it's a good thing, too*. Oxford University Press.
- [Floridi, 2014] Floridi, L. (2014). *The fourth revolution: How the infosphere is reshaping human reality*. OUP Oxford.
- [Floridi, 2015] Floridi, L. (2015). *The onlife manifesto: Being human in a hyperconnected era*. Springer Nature.
- [Floridi, 2021] Floridi, L. (2021). Trump, Parler, and Regulating the Infosphere as Our Commons. *Philosophy & Technology*, pages 1–5.
- [Manders-Huits, 2010] Manders-Huits, N. (2010). Practical versus moral identities in identity management. *Ethics and information technology*, 12(1):43–55.
- [Mill, 1859] Mill, J. S. (1859). On liberty.
- [Moor, 1985] Moor, J. H. (1985). What is computer ethics? *Metaphilosophy*, 16(4):266–275.
- [Popper, 1945] Popper, K. R. (1945). The open society and its enemies. *Vol. I, The Age of Plato*.
- [Stewart, 2021] Stewart, E. (2021). Detecting Fake News: Two Problems for Content Moderation. *Philosophy & Technology*.
- [Vallor, 2021] Vallor, S. (2021). Social Networking and Ethics. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, fall 2021 edition.
- [van Mill, 2021] van Mill, D. (2021). Freedom of Speech. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2021 edition.
- [Weckert, 2000] Weckert, J. (2000). What is so bad about Internet content regulation? *Ethics and Information Technology*, 2(2):105–111.

⁴<https://plato.stanford.edu>