

Dionysia Mylonaki, artist and lecturer in Creative Media and Digital Cultures at the University of Hertfordshire.

Panagiotis Tigas, Engineer and Researcher in Artificial Intelligence

Computational censorship in the form of fake news and toxic comments regulation is a subject that comes up quite often in the public discourse, as a result of the volatile political circumstances on a global scale and due to the unquestionable impact of journalism on these circumstances. Public attention has been directed to the role of mainstream and other media in the formation of public opinion, either in the form of articles or in the form of user-generated comments. The purpose is to analyse and allow a deeper understanding of a project that is under development, namely, computational-censorship and to show that algorithmic regulation is not a solution, but rather another layer to a more fundamental problem.

This article examines the implications of developing ML/AI which aims to regulate the internet and we attempt to allow a glimpse into the technical aspect of the problem as a way to back arguments that could be rejected by the ML/AI research community as “non-pragmatic”. Finally, it aims to highlight the absurdity of the current approach to research in this area, which is the exact opposite of the rationalism that the field claims to be embracing.

Ventures such as Google, Twitter and Facebook have revealed their intention to deal with deception (whatever this means) in the online realm while encouraging conversation (Greenberg). A case study is the project Conversation AI by Google, which has been working on Perspective, an API that uses machine learning models to assess the “toxicity” of comments online and label them. Google has already responded to accusations that the aim of the project is not to censor the internet but rather the exact opposite, namely to tackle censorship (Greenberg). But this paradoxical intervention is something that stems from the no-platformism that has re-emerged in the public discourse and which is very central to the rhetoric that underpins regulation. No-platformism online will be discussed as a form of coding and reinforcing legitimate behaviours, as well as the absurdity of the commons being regulated by the markets. However, it is worth starting with the technical obscurity of the problem that has opened the door to the illusion of a solution.

Taming the Wicked

Social problems are not strictly definable and therefore not solvable by machines and algorithms, a common property of what has been classified as “wicked” problems since 1973 (Rittel and Webber). It is worth tying everything back to the definition of Machine Learning / Artificial Intelligence (quite minimal but still accurate), as the scientific field of predictions and extrapolations from data sets (Poole and Mackworth). For an ML/AI problem to be solved, a dataset containing annotated data is needed. Additionally, a formal method of measuring the error between the predicted and actual value is required; this formal method works as a mathematical description of the problem in question. The main issue is that this requires a close-ended and well-defined problem which, in the case of fact-checking, cannot exist. In Dilemmas in a general theory of planning, the authors have classified the problems into two categories, as tame and wicked (Rittel and Webber). Howard Collins has offered a different reading to this classification by shifting attention to actions; polymorphic and mimeomorphic actions differ in the sense that the former draw from one’s understanding of society (and what society means) in comparison with mimeomorphic actions which tend to not show any variation; thus, machines are defined as the entities that do not engage with polymorphic actions (Collins). This is not a matter of how advanced the field of ML/AI is to this day or a given day in the future but rather a matter of formulating a societal issue that is not meant to be formulated.

AI research has stemmed away from its mothership of cognitive science and philosophy. It has become a playground of engineers with silicon valley flavoured “solutionism” who sometimes attempt to use ML/AI “...to fix problems that don’t exist, or for which there is no technological solution, or for which a technological solution will exacerbate existing problems and fail to address underlying issues...”, according to Privacy International (Kaltheuner and Polatin – Reuben 3). Students land AI research opportunities, in a potentially powerful field, with a good understanding of the STEM subjects but with little background knowledge in Humanities, which offer tools for approaching and framing ambiguities. However, this is not a recent phenomenon and adding to our arguments regarding rationality Philip Agre writes in 1997:

As an AI practitioner already well immersed in the literature, I had incorporated the field’s taste for technical formalization so thoroughly into my own cognitive style that I literally could not read the literatures of non-technical fields at

anything beyond a popular level. The problem was not exactly that I could not understand the vocabulary, but that I insisted on trying to read everything as a narration of the workings of a mechanism (Agre 145).

What is important to note is the lack of diversity in the approach of AI research in fields that are non-technical and ambiguous; for instance, treating the problem of fake news as an engineering problem hides fallacies that might be the subject of research and debate within Humanities. The AI Now 2017 report calls for participation from disciplines beyond computer science and engineering not only as an attempt to ensure input plurality in AI research but also as a methodology that distributes decision-making power (Campolo et al. 2).

In our case, attempting to define the problem of fact checking as a classification problem is prone to fallacies; it requires a definition of the term “fact” that admits a true or false label and, although this might be the case with facts to a great extent (e.g. “the earth is flat”), there are facts that are far from easy to categorise as true or false (e.g. “Islamic State is the consequence of...”) and that would require a thorough study of the epistemology of facts. Similarly, labelling toxic comments and hate speech is equally problematic, politically and consequently, technically. Arguably, the reality is not composed strictly of facts; an automated process in journalism, for instance, would not lack the critical eye required but worse, would undermine the plurality required for journalism to qualify as journalism. On the other hand, crowdsourcing (e.g. Wikipedia) seems to have more appropriate mechanisms embedded and motivations of keeping a bias-free content (bias-free would not necessarily mean free of bias but free of hidden bias; for instance, a debate works as a bias reduction mechanism by exposing the biases).

To return to the Rittel and Webber classification, it is easy to see that the fake news, as well as the toxic comments challenge, fall into the Wicked Problem category (161-167):

1. There is no definitive formulation. The information needed to understand the problem depends upon one’s idea for solving it. Formulating a wicked problem is the problem.
2. There is no stopping rule. Because solving the problem is identical to understanding it, there are no criteria for sufficient understanding and therefore completion.
3. Solutions are not true or false, but good or bad. Many parties may make

(different) judgments about the goodness of the solution. (See Plotzen’s Caliph paper.)

4. There is no test of the solution. Any solution generates waves of consequences that propagate forever.
5. Every solution is “one-shot” — there is no opportunity to learn by trial and error. Every solution leaves traces that cannot be undone. You can’t build a freeway to test if it works.
6. No enumerable set of solutions.
7. Every wicked problem is unique.
8. Every wicked problem is a symptom of another problem.
9. Wicked problems can be explained in many ways. My interpretation is that this is the dual of “no right solution” — no obvious cause.
10. The planner has no right to be wrong. The planner is responsible for the wellbeing of many; there is no such thing as hypotheses that can be proposed, tested, and refuted.

Therefore, the definition of the problem, as well as the extensive research in algorithmic biases, reveals, at best, the fact that the area is known of being prone to biases. Kate Crawford, who has studied the social implications of Machine Learning for years, in the NIPS 2017 conference asked “what if bias is always going to be a problem?” allowing a glimpse into the precarious mechanisms of classification (Crawford). If this is the case, we can only assume that there has been an effort by those who promote AI regulation as a solution to brand a bias-prone service as the only rational, legitimate, universal truth provider, on the basis of the fact that Artificial Intelligence is a black box to the majority.

“Senator, we run ads” or the Revenue

Paradox[1]

The second point that makes the endeavour questionable is the fact that, paradoxically, companies involved in the advertiser/consumer loop (Google and Facebook) are the ones promising to tackle the problem. Considering the economics of fact checking, it is true that automated tools for this task will reduce the cost of media companies, however, the role of human fact-checkers has been reduced (Stencel) without having being replaced by robots. This is not particularly surprising in the attention-hungry economy of the internet; emotionally charged articles are usually more profitable. The so-called data-driven development has

become the dominant paradigm in the computational Ads space, applying A/B testing [2]. Such data-driven corporations are tuning their algorithms using real-time analytics following mostly the metrics related to user engagement and revenue. To elaborate on A/B testing, during the user experimentation phase, engineers observe certain metrics through data collection. These metrics can track click-through rates (number of times a user clicked when she/he encountered an Ad), revenue per impression and other behaviours that work as an approximation of the intended behaviour (in this case, the goal is to direct the user to click on ads and contribute to the revenue of the search company). If the “treatment” (the new algorithm to be tested) improves the metrics for its subset of users compared to the control group (the group subjected to the existing algorithm), it gets deployed and this results in an update of the search / Ad recommendation algorithm. The overall debate surrounding the ethics of user experimentation in AI such as A/B testing has been presented before by other authors (Bird et al.) but, in our case, the irony resides in the fact that attempts to change the nature of the algorithm have to be compatible with the revenue model and to, therefore, maximise content profit. Will these companies deny the profits of click-baiting content when they actively turn any user less capable of resisting clicking on profitable content? An automated fact checker could harm the user engagement/revenue metrics and therefore would not be appealing to the investors who are the ultimate decision-makers.

Consequently, the ones who created the problem in the first place are unlikely to resolve it, as this is not part of their business model. A more likely scenario in this direction would be to see these ventures defining the “fake” and the “toxic” in accordance with the needs of the profit-making machine.

Indeed, the following patent, which seems to be owned by Microsoft currently (April 2018) and previously held by LinkedIn, is a good example of how this paradox passes unnoticed:

The fact checking system will provide users with vastly increased knowledge, limit the dissemination of misleading or incorrect information, provide increased revenue streams for content providers, increase advertising opportunities, and support many other advantages (Myslinski).

It is worth noting that the patent even attempts to define and formulate hypocrisy in an effort to identify and flag hypocritical statements (ibid).

But what is it that makes such companies invest in politics and regulation of public hysteria when they successfully capitalise on this hysteria? A possible explanation, which is more thoroughly discussed in one of the following sections, is that they attempt to come to terms with governments (see also Greenwald; BBC). Other authors such as Christian Fuchs put the emphasis on the media altogether and interpret this moral panic revolving around popular culture as an attempt to distract from the factors that gave rise to social unrest in the first place. On the other hand, endeavours such as Conversation AI and, more importantly, <https://jigsaw.google.com> demonstrate some active interest in regulating the political commons. And worryingly enough, Google attempts to identify radicalisation and propaganda (Jigsaw), two notions that are very central to state terrorism and colonisation.

This brings up the issue of self-censorship. It might not be important whether these companies will formulate a technical approach to the “problem”, as the users themselves will offer a “solution” which will be perfectly aligned with the status quo; several studies after the NSA/PRISM revelations showed that there has been a chilling effect on search behaviour and what we read and write online (Marthews and Tucker; PEN American Center) and to tie everything back to the above paradox, it seems that where self-censorship exists, it leads to damage to the profitability of internet firms (Penney).

A final point relevant to profit-making to be considered is user experience; creating intelligent systems for regulation presupposes that the behaviour of counterfeiters is stationary (does not change over time). In other systems, such as CAPTCHA or Ad Blockers, there is an adversarial relationship between the counterfeiter and the regulating entity, forcing the counterfeiter to adapt and therefore evolve. It is true, though, that the same applies to the regulating entity, however, what is left out of this equation is the user who has become more and more “protected” and suppressed. This raises the question of whether fake news and hate speech are going to remain high on the agenda, under the threat of creating an absurd and intimidating experience for the user.

West-centric, Liberalism Driven

The third point that is indicative of the research values that underpin AI/ML development is rooted in the power relations that are reproduced by the algorithms. Although biased algorithms come to the surface regularly, the

representatives of big ventures are comfortable with public apologies as they usually respond that their approach is liable to the ideas circulated online and the state of the internet as a whole (Thompson) and, worst case, the unconscious biases of the engineers. In this way, what is rarely questioned is the agenda of their research methodology and where the right to exercise authority stems from.

The above argument regarding the source of biases which attempts to pinpoint the general public consciousness as the root cause is tenuous, as it overlooks the limited breadth of the internet base and the factors behind the digital divide. We already know that the internet base is asymmetrical as certain populations and classes are underrepresented (Hopf and Picot; Goldfarb and Prince 2-15). The demographics of data in our case are infused with western rationalism, showing that they are west-centric and liberalism driven and the below screenshots of Perspective API are very explicit in this sense; western leaders' names seem to be protected from toxic comments (Figure 1) while names of other leaders do not (Figure 2). This is not surprising if we consider the new direction of racism, which is exercised on a cultural basis (Hardt and Negri 190-195). Even the fact that the term “fake news” became popularised and associated with the 2016 U.S. elections (Figure 3) and a series of European voting processes with discussions around external interventions in the background shows that the goal was not to open a conversation around values in journalism but rather to start tackling a problem that threatens the integrity of individual democracies [3]. Therefore, there is no framework that could potentially legitimise computational censorship universally and for all classes.

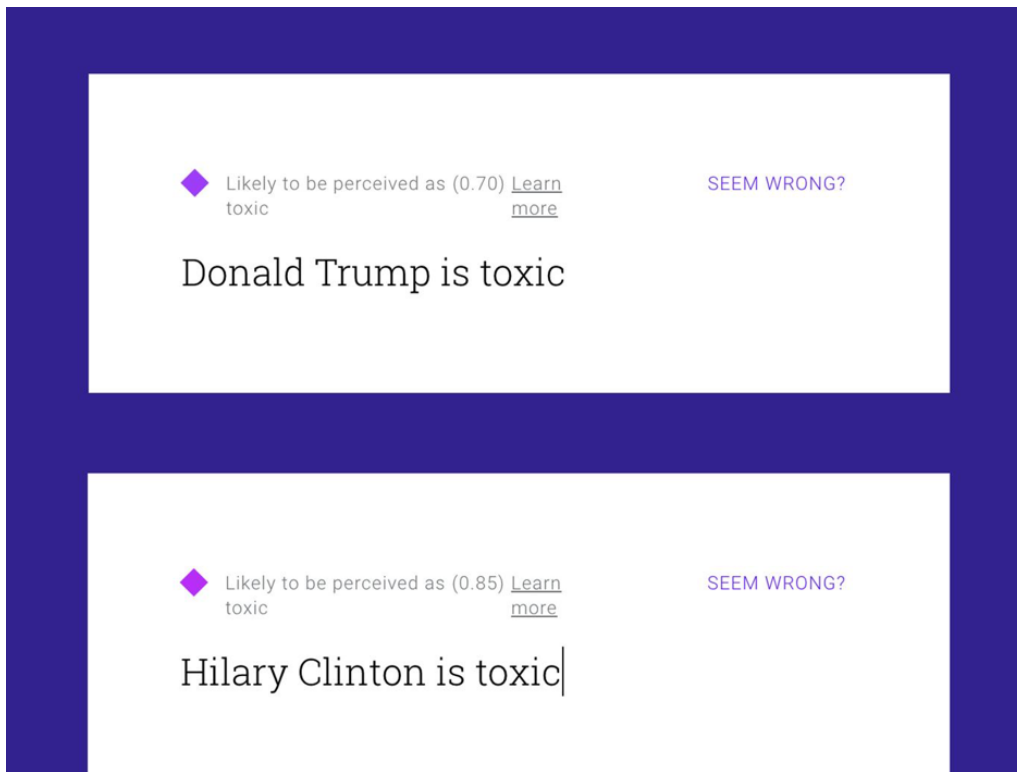


Figure 1: Screenshot from perspectiveapi.com by Conversation AI, Google. Image by the authors taken in January 2018.

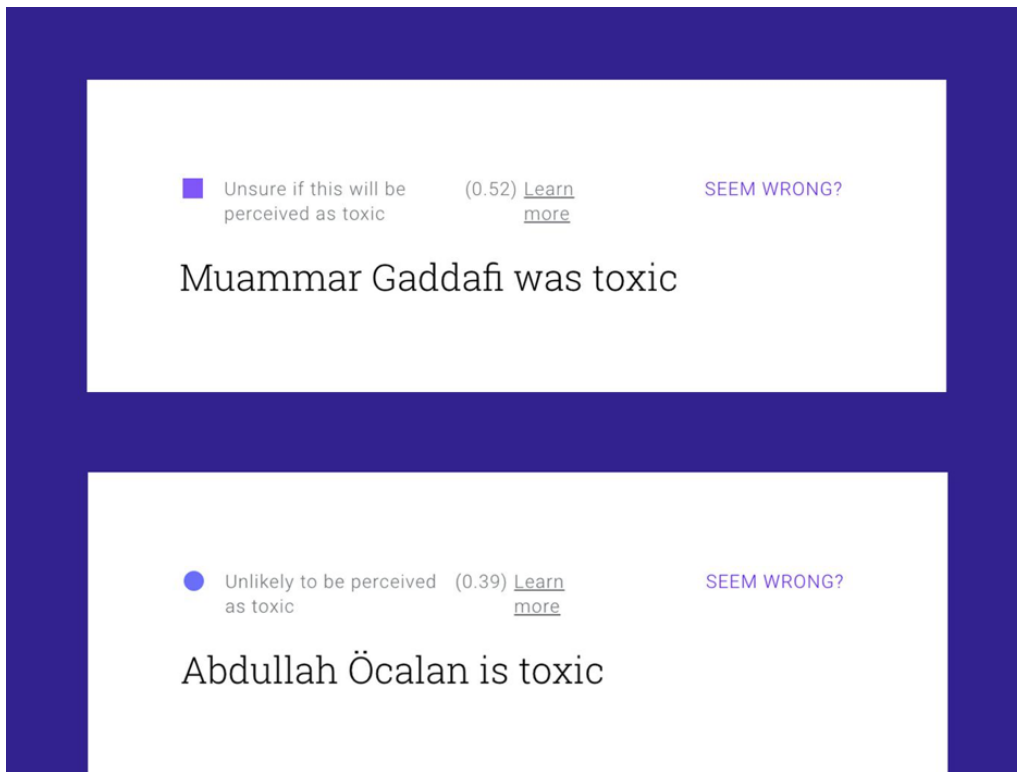


Figure 2: Screenshot from perspectiveapi.com by Conversation AI, Google. Image by the authors taken in January 2018.

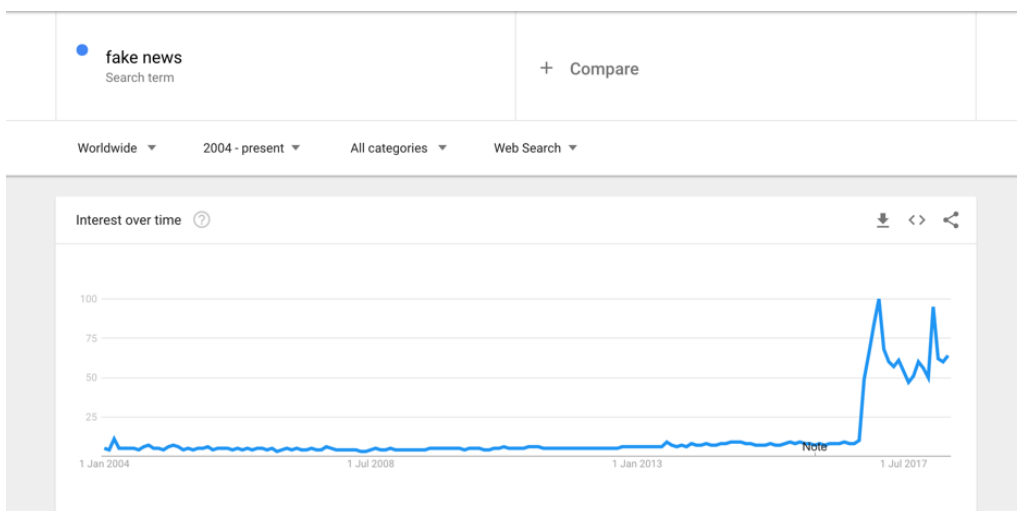


Figure 3: The popularity of the term “fake news” on Google Trends. Image by the authors.

In other words, instead of discussing the biases of algorithms, which, in fact, does

not question but endorses techno-determinism, we should start discussing the neoliberal agenda of algorithms. This is not a question of how we develop algorithms but rather how we conduct research. Focusing on biases behind algorithms depoliticise the conversation, giving the impression that this is an issue either at the level of the engineer or at the level of the user. It is the research agenda that is pro-capital biased.

One manifestation of the pro-capital research agenda is the “Move Fast and Break Things” strategy (as Mark Zuckerberg was once quoted) which is asked to be embraced by researchers and engineers and which demonstrates a quantitative rather than a qualitative and socially accountable approach (see also Taplin).

“Illegitimate” Culture

Although the above points seem to address the technical aspect of the problem, in reality, the described fallacies stem from a combination of the social, the political and the technical. The question is not whether computational censorship and regulation as a solution are adequate and efficient but the real question is, a solution to what and efficient for whom.

In the first section we mentioned that the technologies in question are not only far from being a solution, but, in fact, they add another layer to a fundamental socio-political issue. New regulation technologies need to be examined further in terms of how their intervention (that is the act of aggressively suggesting who and what will be considered as toxic and propagandistic) is constructed and how it relates to the current political landscape.

To start with the former, Jack M. Balkin analyses the anatomy of the “new school speech regulation”; this consists of “the Internet backbone, cloud services, the international domain name system (“DNS”), Internet service providers, web hosting services, social media platforms, and search engines” as well as payment systems and intermediaries. He concludes that all three structures which underpin this “new school speech regulation” revolve around indirect forms of censorship. These are collateral censorship, which aims at silencing an individual or organisation by regulating a facilitating entity, public/private cooptation, which aims at public speech via the appropriation of developed infrastructure by the state, either through direct pressure to corporations or jawboning and finally, private governance by infrastructure owners which appears to be legitimate, not

only due to the pressure by nation-states but also due to the pressure by a number of end users themselves (Balkin, “Free Speech”).

This exact indirect interaction is what makes regulation paternalistic, in the sense that it removes any connotation of suppression which has been connected with authoritarian regimes. Thus, when ventures such as Google and Facebook are taking over the role of the moderator (for reasons and in ways discussed above), declaring that they aim to hold back hate speech and fake news, they make sure that the project is communicated not as an act of submission to pressure but as a form of activism, where algorithms will reverse the deteriorating political and economic circumstances.

This exact solutionism underpins power relations and hierarchies; as Evgeny Morozov writes in *Net Delusion*, the quick fix, “taming the wicked” approach makes it tempting to apply quick fixes “more aggressively and indiscriminately” since it’s a relatively cheap approach to social engineering (303). Morozov, too, referring to the Rittel and Webber classification, questions the ability of any formulated approach to wicked problems to produce universally valid solutions (308) [4]. Indeed, western media problems might not be the same as media problems anywhere else, so, there is no such thing as one solution that applies to every environment.

But this exact enforced solution “aggressively and indiscriminately” creates dynamics that are not meant to be confined to the online realm as the offline has become inseparable from the online and the technological when it comes to social life. Thus, beyond decoding the channels of algorithmic regulation, we need to ask who exactly it is primarily that will experience the workings of power relations online and consequently offline. Kroker and Weinstein in their book *Data Trash* elaborate on the different classes that we encounter in the “technotopia”, with the dominant one being what they call the “virtual class”. According to the authors, the virtual class is that which is determined to protect technotopia, excluding any discussion and perspective that challenges and questions “the fully realized technological society”. This class acts against “economic justice” and “democratic discourse”, instituting a cyber-authoritarianism (Kroker and Weinstein 4-8).

Their theory is a possible approach to understanding how classes are being regulated by incognito algorithms, with one’s public presence (be it offline or online) being approved or disapproved. But it raises the question of what it means for specific classes not to be approved by these algorithms in the public sphere in

these volatile circumstances, in a moment when citizens get less and less access to wealth, wellbeing and education. In the case of the AI/ML regulation technologies that are designed to detect anger, the suppressed are not only those who are already underrepresented (as mentioned earlier) but also those who are too angry to submit and show trust to the establishment and mainstream voices of the virtual class that these regulation mechanisms represent.

The answer is again offered by Kroker and Weinstein who speak about retro-fascism, “the reaction of a body that has been humiliated and marginalized by the digitalization of every communicative and social form of exchange. This reaction assumes the aspects of demented aggressive behaviors – demented, because intelligence has been entirely subsumed and absorbed under the abstract machine of info-production” (Berardi and Mecchia). In this case, retro-fascism (or simply fascism) is what occurs where a big part of the population becomes intimidated by the virtual class, as well as by an invisible intelligent entity and where modes of participation in the commons lie beyond the control of citizens but are up to researchers/engineers working for very powerful corporations.

Much of this conversation is happening in the spirit of no-platforming that has reemerged in the face of this exact volatility and the rise of the far-right but Judith Butler discussed “excitable speech”, hate speech regulations and, in some ways, no-platforming two decades ago, with many of her ideas being applicable in an online context. Her arguments are certainly not one-dimensional but she suggests, in a way, that it is absurd to attempt to regulate speech when the “universally” accepted institution “is constituted through racist exclusions” which are there to assure its stability and confirm its legitimacy, as Antonio Negri and Michael Hardt would probably add (124-129). But to return to Judith Butler, she argues that there’s something more fundamental in hate speech than the right to speak itself and this is the instituting mechanisms that generate it, hence the irrationality of the attempt to regulate it (90). In other words, hate speech is only a symptom of institutionalised exclusion and computational censorship (similarly to censorship of any kind) aims to beautify the internet, concealing only the symptoms of the unstable global circumstances.

Indeed, censorship as an idea has been connected with considering as illegitimate anything that threatens the unity and integrity of a body, hence its association with the state. But in the online realm, where there is no homogeneity to be protected against “external” factors, what is it that is threatened by fake news

and toxic comments? The answer to this question might be the one discussed above, i.e. unity within this or that state in the face of social unrest and a possible far-right outburst. But beyond that, what is being protected is the integrity of the neoliberal establishment on a global scale [5].

Indeed, in July 2017, the World Socialist Web Site (WSWS) reported that Google’s algorithmic updates that were aiming to make it harder for “fake news” and “conspiracy theories” to appear, dramatically reduced traffic to left-wing and anti-war websites, as well as to rights organizations. The relatively long list includes Wikileaks, Truthout, American Civil Liberties Union, Amnesty International and WSWS itself, among others (Damon and Niemuth). Google justified the action taken by explaining that their goal was to prevent “upsetting user experiences”, which reveals the implications of “legitimate” and “illegitimate” political opinion online. But beyond that, although curating information in this way is not synonymous with removing information, it raises questions about whether there is practically any internet outside of Google.

Neoliberalism as an Algorithm

Claiming that neoliberalism is an algorithm might be an extravagant statement to make, in the sense that we can hardly see it as a mathematical construct and it might oversimplify a long process of institution and a more recent process that David Harvey called a counterrevolution and “a political project to re-establish the conditions for capital accumulation” (19). But its similarity to an algorithm lies in the fact that, as an ever-developing project, it relies on processes that aim to profit maximisation through competition and natural selection (survival of the fittest). Therefore, we can hardly say that the above fallacies challenge the actual logic embedded in such projects; questioning the rationality of this process altogether would mean questioning the efficiency of the profit maximisation process for the elites, something that we know is unquestionable. In other words, although the above paradoxes attempt to question its rationality, in reality, they do not challenge its *raison d’être*. Despite the fallacies (and because of the fallacies) demographics of data and capital can spread the western “civilised values” online, fake news can be less obviously fake and socially complex problems can be formulated, being reduced to the level of technology without affecting profit-making.

It is worth noting that authors such as Lawrence Lessig, Frank Pasquale and Jack

M. Balkin see the law as the possible catalyst to disrupt “omniscience”, in combination with public demand for transparency and accountability (Lessig; Pasquale; Balkin, “Three Laws”). But thinking of law as a tool against regulation might be paradoxical, especially when both are the product of the same “algorithm”. However, Lessig clarifies that the question is not “regulation” or “no regulation” as the code is regulative by nature. He suggests promoting decentralisation but at the same time, he urges us to think what kind of private interests step in when the state steps aside (ibid.).

Exposing the research agendas is not enough. This is not to say that neoliberalism must be thought as a determined, fatal condition from which there is no escape. As a final and more positive note, we can say that this condition is susceptible to the scholarship of individual researchers. Following the example of other STEM-related fields, such as that of Development Studies, the AI/ML field can be enhanced with decolonising research methodologies, teaching how they interplay with different classes, territories and political landscapes, introducing not only elements of sociology but also the political and the anthropological. This new scholarship would not take for granted and disseminate the over-productive, western, liberal rationalism, as the only principle that should underpin research. Without a more “instituted” discussion, the neoliberal algorithm proves capable of presenting itself as thoroughly researched, universally legitimate and democratic to the public consciousness, thanks to its patriarchal and patronising underlying mechanisms that are perfectly aligned with the values of a several-thousand-year instituting society.

Notes

[1] The title refers to the response of Mark Zuckerberg when he was asked by Sen. Orrin Hatch about his business model (Liao).

[2] A/B testing refers to a process in product development where users are shown two different versions of a given service with each user accessing only one version in order to determine which version improves the metrics in question (Kaufmann, Cappe, and Garivier 461-481).

[3] U.S. Senator Lindsey Graham, Chairman of the Subcommittee on Crime and Terrorism, declared in October 2017 that manipulation of social media “by terrorist organizations and foreign governments is one of the greatest challenges

to American democracy”, as well as a threat to the U.S. national security. The subcommittee invited Facebook, Twitter and Google representatives to testify.

[4] Of course, what Evgeny Morozov had in mind was grassroots movements, rather than top-down solutionism but the same limitations and precariousness apply to both approaches.

[5] Here, the idea of a body that seeks to protect its unity on a global scale as it is manifested through the transnational moral panic against fake news and toxic language, is developed building on what Arjun Appadurai has defined as “Ideocide”, in his book *Fear of Small Numbers: An Essay on the Geography of Anger*; the phenomenon “whereby whole peoples, countries, and ways of life are regarded as noxious and outside the circle of humanity”, targeting “‘internal’ minorities”, “whole ideologies, large regions and ways of life” (117).

Works cited

Agre, Philip E. “Toward A Critical Technical Practice: Lessons Learned In Trying To Reform AI”. *Social Science, Technical Systems And Cooperative Work: Beyond The Great Divide*, Geof Bowker et al, 1st ed., Taylor & Francis Group, New York, 1997, p. 145. Print.

Appadurai, Arjun. *Fear of small numbers: an essay on the geography of anger*. Duke Univ. Press, 2006. Print.

Balkin, Jack M. “Free Speech In The Algorithmic Society: Big Data, Private Governance, And New School Speech Regulation”. *SSRN Electronic Journal*, 2017, pp. 27-39. Print.

—. “The Three Laws of Robotics in the Age of Big Data”. *Ohio State Law Journal*, vol. 78, 2017. Web <<https://ssrn.com/abstract=2890965>>.

Berardi, Franco, and Giuseppina Mecchia. “Technology And Knowledge In A Universe Of Indetermination”. *Substance*36.1(2007), pp. 57-74. Print.

Bird, Sarah, et al. “Exploring or Exploiting? Social and Ethical Implications of Autonomous Experimentation in AI.” *Workshop on Fairness, Accountability, and Transparency in Machine Learning (FAT-ML)*, 2 Oct. 2016. Web <<https://ssrn.com/abstract=2846909>>.

Butler, Judith. *Excitable Speech*. New York, Routledge, 1996. Print.

Campolo, Alex, et al. AI Now 2017 Report. AI Now, 2017. Web
<ainowinstitute.org/AI_Now_2017_Report.pdf>.

Collins, Howard. The Shape of Actions: What Humans and Machines can do. MIT Press, 1999. Print.

Crawford, Kate. “The Trouble with Bias.” NIPS 2017 Keynote. NIPS 2017, Dec. 2017, Long Beach , CA. Web
<https://www.youtube.com/watch?v=fMym_BKWQzk>.

Damon, Andre, and Niles Niemuth . “New Google algorithm restricts access to left-Wing, progressive web sites.” WSWS, 27 July 2017. Web
<www.wsws.org/en/articles/2017/07/27/goog-j27.html>.

Fuchs, Christian. “Social Media And The Uk Riots: Twitter Mobs, Facebook Mobs, Blackberrymobs And The Structural Violence Of Neoliberalism”. Information - Society - Technology & Media, 2011. Web<<http://fuchs.uti.at/667/>>.

“Germany starts enforcing hate speech law.” BBC, 1 Jan. 2018. Web
<www.bbc.co.uk/news/technology-42510868>.

Goldfarb, Avi, and Jeff Prince. “Internet Adoption and Usage Patterns are Different: Implications for the Digital Divide.” Information Economics and Policy 20.1(2008), pp. 2-15. Print.

Graham, Lindsey. “Graham: Manipulation Of Social Media Sites ‘One Of The Greatest Challenges’ To American Democracy.” United States Senator Lindsey Graham, 31 October 2017. Web
<<https://www.lgraham.senate.gov/public/index.cfm/press-releases?ID=212164DA-EF6F-42E4-91E6-F1544014D134>>.

Greenberg, Andy. “Inside Google’s Internet Justice League and Its AI-Powered War on Trolls.” WIRED, 19 September 2016. Web
<<https://www.wired.com/2016/09/inside-googles-internet-justice-league-ai-powered-war-trolls/>>.

Greenwald, Glenn. “Facebook Says It Is Deleting Accounts At The Direction Of The U.S. And Israeli Governments.” The Intercept, 2017. Web
<<https://theintercept.com/2017/12/30/facebook-says-it-is-deleting-accounts-at-the-direction-of-the-u-s-and-israeli-governments/>>.

Hardt, Michael, and Antonio Negri. *Empire*. Harvard University Press, 2000, pp. 124-195. Print.

Harvey, David. *A Brief History of Neoliberalism*. Oxford University Press, Oxford, 2007. Print.

Hopf, Stefan, and Arnold Picot. “Are Users All the Same? – A Comparative International Analysis of Digital Technology Adoption.” *Homo Connectus*, edited by Keuper F, M. Schomann and L. Sikora L. Wiesbaden: Springer, 2018, pp. 103-119. Print.

“jigsaw.” Google. Web <jigsaw.google.com/challenges/>.

Kaltheuner, Frederike, and Dana Polatin – Reuben. *Submission Of Evidence To The House Of Lords Select Committee On Artificial Intelligence*. Privacy International, 6 September 2017, p. 3. Web <<https://privacyinternational.org/sites/default/files/Submission%20of%20evidence%20to%20the%20House%20of%20Lords%20Select%20Committee%20on%20Artificial%20Intelligence%20-%20Privacy%20International.pdf>>.

Kaufmann, Emilie, Olivier Cappe, and Aurelien Garivier. “On the Complexity of A/B Testing.” *Journal of Machine Learning Research*, vol. 35, 2014, pp. 461-481. Print.

Kroker, Arthur, and Michael A Weinstein. *Data Trash*. Apogeo, 1996. Print.

Lessig, Lawrence. “Code is Law: On Liberty in Cyberspace”, *Harvard Magazine*, 1 January 2000. Web <<https://harvardmagazine.com/2000/01/code-is-law-html>>.

Liao, Shannon. “11 Weird And Awkward Moments From Two Days Of Mark Zuckerberg’S Congressional Hearing”. *The Verge*, 2018. Web <<https://www.theverge.com/2018/4/11/17224184/facebook-mark-zuckerberg-congress-senators>>.

Marthews, Alex, and Catherine Tucker. “Government Surveillance and Internet Search Behavior.” *SSRN Electronic Journal*, 2017. Web <<https://ssrn.com/abstract=2412564>> or <<http://dx.doi.org/10.2139/ssrn.2412564>>.

Morozov, Evgeny. *The Net Delusion*. London, Penguin, 2012. Print.

Myslinski, Lucas J. *Social media fact checking method and system*. U.S. Patent No.

8,458,046. 4 Jun. 2013.

Pasquale, Frank. Black Box Society. Harvard University Press, 2015. Print.

PEN American Center. Chilling Effects: NSA Surveillance Drives U.S. Writers To Self-Censor. PEN American Center, 2013. Web

<https://pen.org/sites/default/files/2014-08-01_Full%20Report_Chilling%20Effects%20w%20Color%20cover-UPDATED.pdf>.

Penney, Jon. “Chilling Effects: Online Surveillance and Wikipedia Use.” Berkeley Technology Law Journal, 31.1(2016): pp. 117. Web

<<https://ssrn.com/abstract=2769645>>.

Poole, David L., and Alan K. Mackworth. Artificial Intelligence: foundations of computational agents. Cambridge University Press, 2010. Print.

Rittel, Horst W. J., and Melvin M. Webber. “Dilemmas In A General Theory Of Planning.” Policy Sciences, vol 4, no. 2, 1973, pp. 155-169. Springer NaturePrint.

Stencel, Mark. “A Big Year For Fact-Checking, But Not For New U.S. Fact-Checkers.” Duke Reporters’ Lab, 13 December 2017. Web

<<https://reporterslab.org/category/fact-checking/>>.

Taplin, Jonathan. Move Fast And Break Things. Macmillan, 2017. Print.

Thompson, Andrew. “Google’s Sentiment Analyzer Thinks Being Gay Is Bad.” Motherboard, 25 October 2017. Web

<https://motherboard.vice.com/en_us/article/j5jmj8/google-artificial-intelligence-bias>.