Facebook Fake Profile Identification: Technical and Ethical Considerations

Pardis Pourghomi, Milan Dordevic
College of Engineering and Technology
American University of the Middle East Egaila,
Kuwait
{ Milan.Dordevic, Pardis.Pourghomi}@aum.edu.kw

Fadi Safieddine
School of Business and Management
Queen Mary University of London
London, UK
F.Safiedinne@qmul.ac.uk

Abstract— **March 2019, Facebook updated its security procedures requesting ID verification for people who wish to advertise or promote political posts of adverts. The announcement received little media coverage even though it is an interesting development in the battle against Fake News. This paper reviews the current literature on different approaches in the battle against the spread of fake news, including the use of computer algorithms, A.I, and introduction of ID checks. Critical to the evaluation is consideration into ID checks as a means to combat the spread of Fake News. To understand the process and how it works, the team undertook a social experiment combined with reflective analysis to understand the impact of ID check policies better when combined with other standards policies of a typical platform. The analysis identifies grave concerns. In a wider context, standardising such policy will leave political activists in countries vulnerable to reprisal from authoritarian regimes. Other impacts include people who use fake names to protect the identity of adopted children or to protect anonymity from abusive partners. The analysis also points to the fact that troll arms could bypass these checks rendering the use of ID checks less effective in the battle to combat fake news.**

Keywords— Misinformation; Fake News; Social Media; Algorithm; Artificial Intelligence, Policies; Ethical impact.

I. INTRODUCTION

Nowadays, millions of users are attracted to social media and its different platforms, which enables users to share information, personal interests, news, etc. easily across the web. Since the majority of the unverified information that is shared across social media has been presented as facts, believing or disbelieving such information has become a major challenge. The source of misinformation spread is normally malicious users who are motivated to influence other users' views or motivated by financial gains. Fake news, dotted images, chain emails, misleading news, spam, out of context videos and images are various forms of misinformation. The need for bringing some credibility for verifying content shared online as well as improving users' experience on social media is vital. Hence there have been multiple attempts to develop means and tools to minimise the spread of misinformation on social media. Thus, allowing misinformation flow freely on social media not only wastes users' time but can also be dangerous [1]. Several approaches have been developed to limit the spread of fake news by allowing users to authenticate it from within their web browsers [2], A.I [3], labelling news [4], and many more.

II. LITERATURE REVIEW

The quality of information shared on social media is becoming more significant as more users on social media platforms search for the news [16]. In this section, we present the work done by social media companies in combatting fake news that is reported in academic literature. These papers have explored proposed approaches and techniques that are based on platform policies, digital algorithms, filter, and Artificial Intelligence (A.I.) in the process of detecting fake news.
In [3], authors claim to have developed a process for automating fake news detection on Twitter. They achieved it by learning how to forecast precision assessments in two dedicated Twitter datasets.

The first data set is a credbank, a crowdsourced dataset of precision assessments for events on Twitter. The second one is called pheme. Pheme represents dataset of potential gossips on Twitter and newspapers assessments of their truthfulness. Such a computerised system is said to have been able to detect fake news in popular Twitter threads. As part of the process, the model includes the use of crowdsourced workers in updating and maintaining the datasets. In their experiments, crowdsourced workers outperform both newspapers assessment and models trained workers. The argument being, crowdsourced workers, provide an easy way to categorise true and false stories on Twitter compared to newspapers. This model, while promising, continue to have two key risks associated with building trust in both the technological element and the human element.

The works of [4] looked at trust in rating fake news to be fake and factual. The authors suggested that online fake news can be opposed to some degree of success when appropriate tags are used. The result of their study found that both "Disputed" and "Rated false" tags modestly decrease trust in sharing false news; an issue previously explored, described and developed in [9] and [10].

However, their results demonstrated that "Rated false" tags, which exactly articulate to users that the news is not true to appear to be more effective at decreasing trust in misinformation than the "Disputed" tags used by the Facebook. When compared to all-purpose notices regarding fabricated news, it was also found to reduce trust in false headlines. However, the effect of these warnings was too small compared. Additionally, all-purpose warnings also decreased trust in factual news and did not improve the effects of the "Rated false" and "Disputed" tags. The authors have listed several limitations, but despite these limitations, their study showed important perceptions into how efforts to stop trust in misinformation on social media could be more effective.

Another approach to combating social media is shown in [6] where authors attribute the spread of fake news to a term called "filter bubbles". Filter bubbles are the digital echo chamber where users see content and posts that agree only with their preexisting beliefs. While social media platform aims to encourage users of the same interests to share ideas and stay online longer, the unintended consequence is that these users end up living in an information 'bubble'. In such an environment, fake news can go unchallenged but further enforced by the agreeability infested in being shared in a 'filter bubble'. The work of [6] concludes that the dissemination of fake news may, at times, not break out of the filter bubble at its point of origin. In fact, the work of [11] shows that the dissemination of fake news is similar to the epidemic and that such stories typically stay inside the same groups.

Research by [7] categorised important problems that make social media platforms role hard in combatting fight fake news. The researchers described a set of intervention models, categorised under the four regulatory modalities. As part of a discussion, the paper looks at the ease with which fake news generators can provide online content that appears to be authentic as well as the financial stakes that platforms have in sharing fake material online. The principles of the 'filter bubble' seems to influence much of the findings. Clients of fake news have a strong disposition to accept the news and thus limited incentive to participate in challenging and validating its content. A key component in making fake news believable is to introduce fibres of truth [7].

The work of [32] confirms the environment that many suspects from a platform that has grown faster than its policies can keep up. In [32], the researchers review the case study of US elections, where a survey of staffers, politicians, and email exchanges shows that much of Google and Facebook's policies are written on the go. In [8] authors present investigation on a platform-based policy in managing and combatting fake news. The work looks at both self-regulated by social networks and government intervened. From one side, the authors describe how direct government intervention may be accused of biases and lack of independence. Concerns raised that government-based policies may impact freedom of speech. Instead, the authors suggest that governments enable an approach [8] dubbed "tort lawsuits alleging". Under this approach, policies would be shaped by lawsuits initiated by people affected or harmed by the impact of fake news. To facilitate this approach, social networks could provide users with indications of source quality that can be merged into the algorithmic rankings of news.
Moreover, these platforms can reduce the personalisation of political information content and allow alternative reporting of news and views. Thus, remove users from filter-imposed bubbles. Methods that underline at present trending news could be redesigned to allow a variety of news. Social networks could control the automated spread of news content by bots and Cyborgs. In the context of this research, Cyborgs are consumers who automatically share news from a set of sources, sometimes without reading them at all [15].

In [16], the authors showed a framework for detection of bots on Twitter. The system is a machine-learning algorithm that can extracts features divided into six different classes. The first and second classes are users and friends' meta-data. Third and fourth classes are network patterns and activity time series. The last two classes are tweet content and sentiment. The system would evaluate patterns, thus developing a framework which is trained on a dataset of bots. The results suggest an accuracy of 0.95 Area Under Curve (AUC). Thus, a good precision but a margin of error will need enhancing.
Moreover, their analysis suggests that user meta-data and content features are the two most important sources of data to detect typical bots on the social platform. Groups recognised by this analysis point principally into different types of bots. However, in several instances, the boundary that separates these groups is not obvious, suggesting further research is needed in this area.

In a similar study, authors in [14] investigated automated behaviour of bots on Twitter. Using a flexible and transparent classification scheme, the authors showed the potential of using linguistic features as a means of classifying automated bot activity on Twitter. As these features do not use the metadata provided by Twitter, the classification scheme presented in [14] is also applicable to other social media platforms. The paper suggests that current algorithm-based approaches by social media giants rely on metadata sets to find automated bot accounts; this is done by using variables such as the time between tweets, number of followers, etc. In the same paper, authors demonstrated a classification scheme that uses the natural language text from non-bot users to deliver a standard for detecting accounts that post automated bot posts.

Despite the many advances in algorithm-based detection of fake news, it seemed inevitable that A.I. is the natural partner in building the knowledge that would adapt to variations by determined fake news generators. In [5], the authors showed an important potential application of A.I. in recognising "fake news" online. Communicating in a network vocabulary, A.I would use "latency" in delays

before a transfer of data begins following an instruction for its transfer to perform analysis. The latency that human fact-checkers brings into fake news dissemination appears to be to take too long to soften the harm done by fake news. Thus [5] suggest faster, reliable and more stable A.I. methods are needed in fighting misinformation online. The aim is to have A.I methods that can categorise news as either genuine or fake based on two approaches. The first approach uses linguistic and semantic analysis of written content to make their identification. The second approach uses sharing patterns and rates to categorise news into rumours and non-rumours. A combination of both helps A.I. identify posts that are factual or fake.

The social networks have attempted some of preventing steps and others as described above. In 2018, authors in [8] reported several actions taken by the social media platform to address the spread of fake news. Twitter blocked some accounts connected to Russian users and users connected to those accounts. Facebook announced a determination to shift its algorithm to check for "quality" in its content curation process. Nevertheless, the social networks have not given sufficient fact for evaluation to the research community nor shared their results for peer review to support their efforts. There exists a huge need that social networks work with academic researchers in assessing the scope of the fake news problem and the design and implementation of actions against the sharing of fake news. There is warning that poorly structured policies intended to combat the spread of fake news are sometimes inadvertently harming those fighting fake news. The authors of [28] highlight the case of StopFake, a Ukranian fact-checking group operating on Facebook. The group saw many of its followers' accounts closed for sharing articles intended to point out fake posts as opposed to sharing of fake posts. However, Facebook's algorithm did not seem to distinguish between those who share fake news and those who share news about fake news and in the process, harmed the activities of the StopFake's group.

After the 2016 U.S. election, Mark Zuckerberg announced "of all the content on Facebook, more than 99% of what people see is authentic. Only a very small content is fake news and hoaxes. Overall, this makes it extremely unlikely hoaxes changed the outcome of this election in one direction or the other" [12]. Evidently, at the time, even heads of social media giants were not well informed of the challenges they are facing.

In [17], authors report that Facebook has rolled out a policy that requires businesses and individuals actively promoting political activities and ads to undergo ID verification. The article suggests the ID check is Facebook's attempt to curb Russian's interference in EU elections and preparation to manage the 2020 USA election. For the policy to work, the social media platform is using algorithms and A.I. to identify potential fake profiles. Alex Schultz, Facebook's Vice President of Analytics, explains in the company's news release that the launch of Facebook's algorithm has successfully closed over 100 million accounts [29].

III. RESEARCH METHODOLOGY

This paper attempts to answer three research questions:

1. How effective is Facebook's algorithm in detecting a fake profile?
2. How effective is Facebook's A.I. in learning from its operations?
3. What is the ethical impact of Facebook's changes that introduces ID checks to its policies?

To answer these questions, the team suggested an experiment to test the capabilities of the social media algorithm in detecting a fake profile involved in sharing fake news. Since no experiment of this kind has ever been conducted before in the literature, there was a need to design a tailored experiment for this purpose. Following a brainstorming session, the team identified the following key characteristics of the experiment:

1. The profile should have the hallmarks of a fake profile with a fake name and lacks bio information.
2. The photo to be picked as a random none human photo.
3. Profile to accept anyone who is suggested by the platform to be a friend and add all suggested profiles.
4. The profile to avoid political posts but join a variety of groups and repost their posts indiscriminately.

Agreeing that the likelihood of platform algorithm detecting the profile as fake, the team suggested additional matching profiles is created with few modifications to test how well the A.I. aspect of the algorithm is able to learn from its first detection.

In the process, the team will be using reflective analysis to study the process of creating, building, and identifying fake news on the user(s) of profiles with fake names.

Reflective analysis, also known as reflexivity, is a research methodology presented in academic literature [18] [19] [20]. Some academic publications treat this analysis approach as an independent approach or part of a wider approach to transformative learning [21].

In [22], the authors discuss reflective research approach where they capitalise on their personal experiences, analyses them and provides findings that can be generalised to a wider set of circumstances.
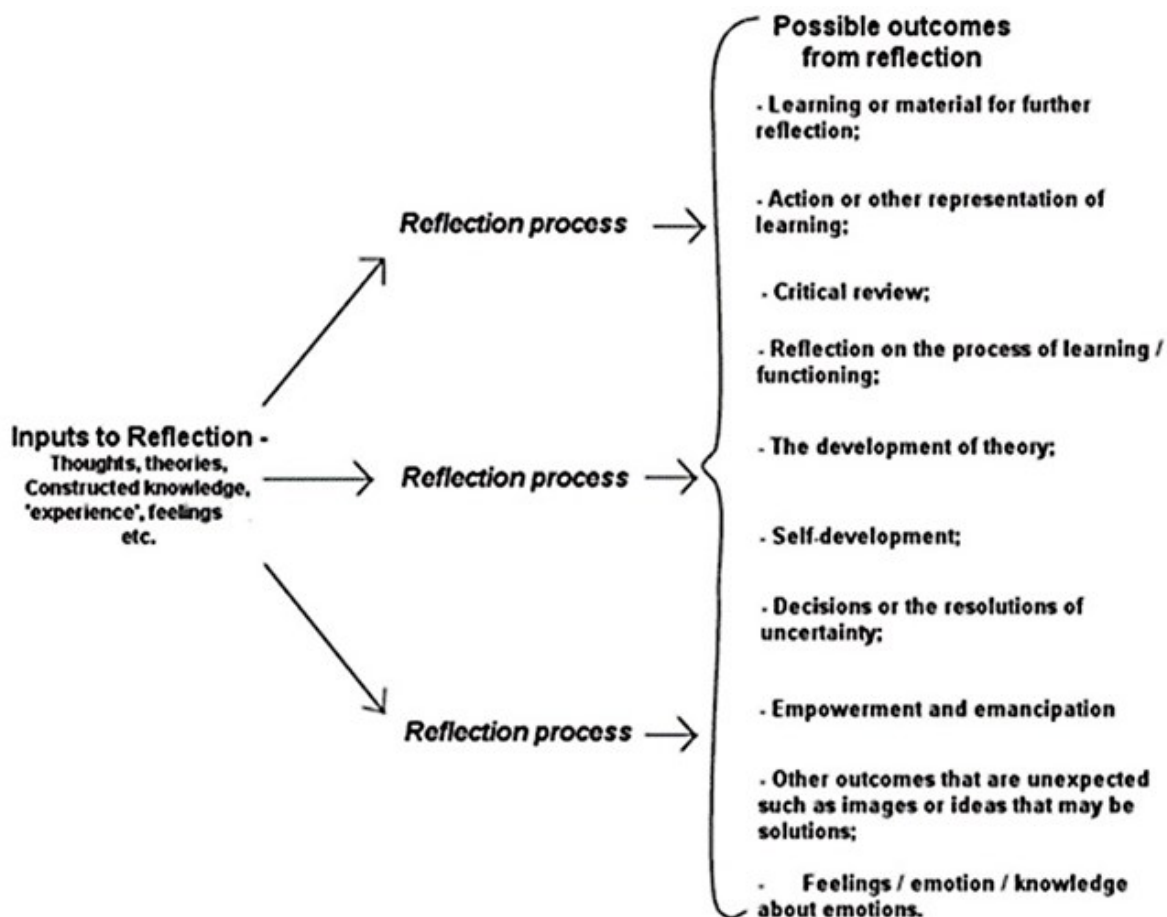
Figure 1. Reflection Process [23]

In the reflection process, academics would use their knowledge that they already have in the form of experience, judgements, concepts, and feelings, coupled with new findings or information to make a reflective analysis of a given situation [23]. In [23], the author suggests that a reflective approach is suitable in cases where there is no obvious or immediate solution, and the input is complicated or unstructured. The author presented reflection as a form of thinking that is represented in an input–outcome model, as shown in Figure 1.

The reflective analysis will help learn from the experiment, critical review shortcomings, resolution of uncertainty with regards to impact, and anticipation of impact. Hence, interviewing and observation will be key inputs in the experiment.

IV. THE EXPERIMENT

In theory, the Facebook algorithm picks up a set of patterns and reports from users; then it requires a suspect profile to validate their details, including their ID and home address. It is most definitely Facebook latest attempt to curb fake news. A summary of the experiment process is provided in figure 2.
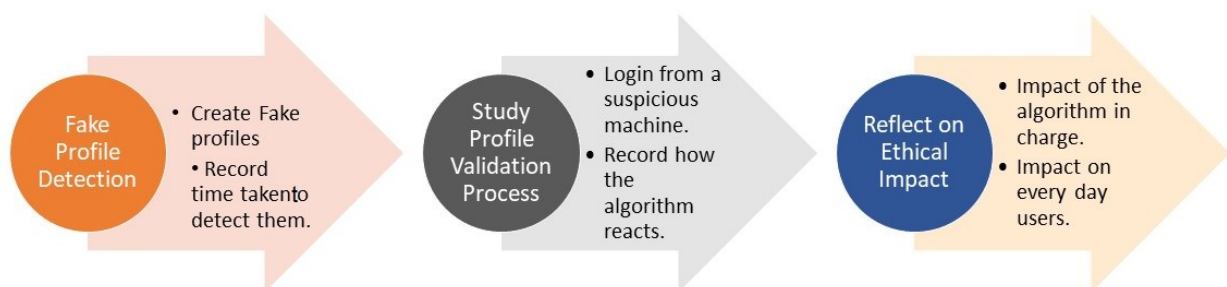
Figure 2. Research experiment process

Stage 1: Fake Profile Detection

To test the system, a single new profile under an assumed Arabic name. The name used is a generic name with random Arabic stock photo of an oriental carpet, and no bio details whatsoever. To trigger the algorithm, the profile rejected the requests to include a mobile number, connect to email to invite friends, and went about adding all friend suggestions put forward by the platform. To further antagonise the algorithm, the profile started almost instantly posted and sharing a variety of liberal and anti-fundamentalist posts. By accepting anyone and everyone who wants to a friend, the profile massed more than 500 friends in less than three weeks. There was no indication that most of the other profiles that linked to our profile looked genuine either.

To maintain a sustainable bot like activities, the team took turns to keep the random posting and engagement of the profile in nothing but post sharing. The team noted concerns that profiles and pages are being closed by the algorithm, thus by week three the algorithm had yet to identify the fake profile we had created. On day 22, the algorithm identified the profile as fake. Initially, the algorithm locked the profile demanding confirmation of a telephone number and photo. The analysis of activities that triggered the lockout was inconclusive. It was what made the algorithm. Several theories emerged including patterns of postings; duration passed where the profile rejected attempts to provide valid bio content, a flood of post sharing done the night before, possible reporting by other users, or combination of these building a pattern identified by the algorithm.

The team recreated the same profile with minor variations of the name, change of email, a similar photo and no bio. Timewise, the algorithm, however, took only 24 hours to lock the profile out. The team was able to narrow reasons why the algorithm got triggered to similar patterns. The profile moved to join the same groups, to share the posts, and to link to the same individuals. Another interesting development this time was the fact the locking out happened while online. The profile was labelled as fake and locked out.

In a final experiment, the team attempted to take some credible steps to mask the profile. Suspecting that the algorithm is likely to be identifying the connection, browser cookies, and other features linked to the machine, the team used a different machine with a different IP connection. Two variables remained the same, a fake name, no bio, and rejecting any attempts to link to friends via emails or mobile phone. Still, the A.I. element of the algorithm had learned of a pattern and locked profile out in under 30 minutes. At this point, there was no longer viable reason to continue experimenting as the A.I. element of the algorithm had learned key aspects of the process and was ready to lock the profile out faster than any time before.

Stage 2: Studying the profile validation process

To further test the algorithm's abilities in identifying machines linked to fake profiles, the team attempted to login to a genuine profile from the first machine used to create fake profiles. One of the team members decided to play the role of a genuine user and to log in using his genuine profile details. From what the algorithm now suspects are a fake serial profile generating machine, Facebook redirected us to a welcoming message requiring additional security checks. The message advised the user to increase security and to include either 2-tier login checks using the Authenticator App or using the registered email. We chose to use the Authenticator App.

To verify the profile as genuine, we elected to apply for the voluntary ID check, as shown in figure 3.

First, the system asked to confirm the home country. Then required a full home address and the system advised that it will cross-check that information. The indication being a UK electoral record check as the processing took a fraction of a minute and it came back with 'Verified'. Should the address come back as invalid, the system would have asked to confirm the address by post whereby the user will have to wait for a letter from Facebook with a code. Finally, the system requested an upload an identification document giving us several options with the key ones being: Driving License, Passport, Birth Certificate, Visa, Voter ID Card, etc. It is evident that these documents are key to any claiming of UK residency. Our user-submitted a scanned copy of a passport; in less than a minute, a small message flashed on the corner of the screen stating the profile has been successfully validated. The user is now able to share and pay political ads and posts, see figure 4.

Figure 3. Facebook ID Checks

Figure 4: Facebook Confirmation of identity message.

Stage 3: Reflection and Ethical Impact

ID checks are a new method the combines algorithm and A.I. to enforced policy-based approach in the battle against the spread of fake news. The policy aims to close down large fake news generators and newsgroups, especially those directed or managed

remotely. Requiring ID checks would make it harder to generate the kind of troll armies (also called troll factories and farms) that some have accused Russia of using to influence election outcomes or undermine democracies. But will it work?

In the reflective analysis, we identified some ethical concerns. Enforcing ID checks will have a detrimental impact on individuals who prefer to remain anonymous when sharing perfectly valid political views. There is most definitely a concern among people in many countries who may want to share political views without risk of reprisal from otherwise authoritarian regimes. Facebook policy, if taken literally, requires all users to use their real names. Technically, Facebook does not allow nicknames, pet names, cute names, and metaphors. Individuals are therefore required to disclose their full names, their real photos, and connect with friends and family; all three activities that can help identify who they are.

Another concern is the security of the documents uploaded to the platform. Facebook suggests that the communication is encrypted, and the ID will be deleted within 30 days of the submission. One user from India who runs a secular page critical of religious authorities and fundamentalism had messaged us on messenger saying he had to do the ID check to protect his page from being closed by the algorithm. His concern, however, was that his details might fall into the wrong hands as his life will be endangered. To avoid his conversation being picked up by the algorithm, the user communicated with us using text written on images. Our analysis suggests that many secular and liberal bloggers in countries where they are not able to speak openly or have a high risk of being targeted will have a problem with this new requirement. They are, after all, much more vulnerable to being reported as fake when they cannot use their real names or faces photos.

It could be argued that asking for ID checks is the cost we must endure to stop the weaponisation of social media for political gains by foreign countries. Our analysis of the process suggests, however, that this approach is open to abuse.

Hackers have, in the past, hacked large stashes of people's ID in the UK and USA. Ironically, some were taken directly from Instagram and Facebook, not once but many times [24]. The Facebook validation process relies on what appears to be a basic identity check. An international and organised troll army will be able to overcome these with some success. A combination of having a list of people, copies of their ID's and their electoral home addresses, and troll army may be able to validate a large volume of profiles. Permutations of profiles may suggest that first attempts take three weeks for the algorithm to identify a profile, and three weeks is a long time to influence an election.

A review on our user's profile list of friends shows several acquaintances who use alternative names or photos for privacy or to protect themselves. For example, two parents who changed their Facebook names on the request of their social workers to maintain the privacy of their adopted son; and one friend running away from an abusive domestic partner and the other was harassed by a stranger online. Both changed their names and used generic photos to mask their identities. Our analysis shows that these are not unique cases.

V. ANALYSIS AND OUTCOMES

In answering the research questions, the experiment demonstrated the following:

1. How effective is Facebook's algorithm in detecting a fake profile?

In the first instance, the platform has demonstrated slow detection in identifying a fake profile that extended for over three weeks. However, the system progressively improved in what appear to be machine language learning. Ultimately, the algorithm has been effective in spotting all three profiles and using a combination of approaches that suggests A.I. an integral part of that process. It is clear that the model looks at the magnitude of metrics in flagging a given profile as fake.

2. How effective is Facebook's A.I. in learning from its operations?

In sequence, the fake profiles took three weeks, then twenty-four hours, and finally thirty minutes to be identified. The A.I. element of the platform demonstrated highly active learning from past patterns that is effective at slowing the activities of prolific fake profile generators. Reports from other users on the platform also confirms this finding.

3. What are the ethical impacts of a platform introducing ID checks policy?

The reflective analysis presented in this paper suggests that the approach would change the fabric of the debate and not always in a positive way. It would impact the privacy of many individuals and exacerbate the risk many political activists take in

managing their political activities online. Where individuals are interested in challenging the norm while protecting their identity, social media platform adopting ID checks will hamper their ability to be effective in their approach.

On technical applicability of the algorithm and A.I, one can conclude from the first two research questions that the Facebook model is promising. The social media platform was able to learn and detect our fake profiles at a rate of three weeks, twenty-four hours, and lastly thirty minutes. Therefore, this experiment backs Facebook's claims [29] that suggest they have been successful at closing many fake profiles. This success needs to be balanced with an approach that does not violates individual rights and penalising and isolating users who prefer to remain anonymous.

Facebook's approach only attempts to fix one problem in a pool of many problems social media platforms have with fake news. Politically oriented Fake News may be a serious concern, but medical advice fake news also kills. Anti-vax posts and fake cancer treatments are rife of social media. Are social media platforms going to be asking people to upload their medical certificates to be allowed to post or share medical advice? Many of these posts are shared by genuine profiles and individuals motivated by tribalistic beliefs.

Breaking news, local news, human stories, historical events posts, quotes, satirical posts, celebrity news, art news, and science news [25] have all been subject to fake news with some having serious consequences. While there is still no evidence that the policy will be effective in slowing down prolific and determined state-sponsored troll armies.

There is a startling warning from several researchers with regards to an inherited risk with having policies that are enforced by algorithms and A.I. Ultimately, individuals or state-sponsored troll armies who are determined to push their style of news are likely to learn what works and what does not work by means of trial and errors. Thus, determined fake news content generators are likely to develop ways to bypass the system's metrics [26] [27]. The success of this policy, combined with other policies introduced to curb the spread of fake news, will be put to the public test in the US presidential election of 2020.

VI. CONCLUDING REMARKS

This paper has been able to demonstrate how policy based on A.I. and software algorithm enforcement methods can be effective in implementing these regulations at a wide scale. Also, the paper's analysis presented serious concerns to the individual rights that need to be considered in the wider context of fighting fake news. The literature study showed that this would not be the first time shortsighted and less then well-evaluated policies failed to account to appropriate reflective analysis on individual rights.

The paper acknowledges that the experiment is the only representative of one experiment with many variables. The ability to control some of these variables proved to be less effective in identifying what triggers an algorithm to identify fake profiles. In our opinion, the combination of fake profile photo, not being selective in building friendships, lack of friendship behaviours such as likes and comments from friends, and therefore not having displayed characteristics of human networking are most likely the control variables of the AI algorithm that could be tested in future research. All these variables are hard to fake as an individual but may be possible to fake as cyber bot group. Facebook and other platforms are not likely to provide confidential data about the way their algorithms operate in identifying fake users.

Future research should consider the long-term impact of A.I and algorithm-based policies. In addition, there should be some comparison on how different platform to attempt to use A.I and algorithm-based policies to combat fake news. As well as considering how those determined to feed social media platforms, fake news will attempt to bypass these policies. More testing could identify further gaps in Facebook and other social media algorithms. This is particularly important as Facebook, and other platforms are increasingly being used as OpenID providers, a told to authenticate user ID and link to other website and applications [33].

Finally, the current policies that attempt to address segments of the fake news problem by removing it are running the risk of being accused of censorship. The platform is failing to listen to academic research into softer human approaches that suggest providing validation tools that would allow users to make that informed decision [30] [31].

REFERENCES

[1] Dordevic, M., Safieddine, F., Masri, W. and Pourghomi, P., 2016, September. Combating Misinformation Online: Identification of Variables and Proof-of-Concept Study. In *Conference on e-Business, e-Services and e-Society* (pp. 442-454). Springer, Cham.

[2] Halimeh, A.A., Pourghomi, P. and Safieddine, F., 2017. The Impact of Facebook's News Fact-Checking on Information Quality (IQ) Shared on Social Media. In *MIT International Conference on Information Quality*.

[3] Buntain, C. and Golbeck, J., 2017, November. Automatically identifying fake news in popular twitter threads. In 2017 *IEEE International Conference on Smart Cloud (SmartCloud)* (pp. 208-215).

[4] Clayton, K., Blair, S., Busam, J.A., Forstner, S., Glance, J., Green, G., Kawata, A., Kovvuri, A., Martin, J., Morgan, E. and Sandhu, M., 2019. Real solutions for fake news? Measuring the effectiveness of general warnings and fact-check tags in reducing belief in false stories on social media. *Political Behavior*, pp.1-23.

[5] Cybenko, A. K., and Cybenko, G. 2018. AI and Fake News. *IEEE Intelligent Systems*, 33(5), 1-5.

[6] DiFranzo, D. and Gloria, M.J.K., 2017. Filter bubbles and fake news. *ACM Crossroads*, 23(3), pp.32-35.

[7] Verstraete, M., Bambauer, D. E., and Bambauer, J. R. 2017. *Identifying and countering fake news.* Andy Black Associate, UK [Online]. [Accessed 25 May 2019]. Available from: https://www.andyblackassociates.co.uk/wp-content/uploads/2015/06/fakenewsfinal.pdf

[8] Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., and Schudson, M. 2018. The science of fake news. *Science*, 359(6380), 1094-1096.

[9] Ecker, U. K., Lewandowsky, S., and Tang, D. T. 2010. Explicit warnings reduce but do not eliminate the continued influence of misinformation. *Memory & Cognition*, 38(8), 1087-1100.

[10] Bolsen, T., and Druckman, J. N. 2015. Counteracting the politicization of science. *Journal of Communication*, 65(5), 745-769.

[11] Jin, F., Dougherty, E., Saraf, P., Cao, Y., and Ramakrishnan, N. 2013. Epidemiological modeling of news and rumors on twitter. In *Proceedings of the 7th Workshop on Social Network Mining and Analysis* (p. 8). ACM.

[12] Kokalitcheva, K. Mark Zuckerberg says fake news on Facebook affecting the election is a 'crazy idea.' *Fortune (Nov. 11, 2016)*; [Online]. [Accessed 23 May 2019]. Available from: http://fortune.com/2016/11/11/facebook-election-fakenews-mark-zuckerberg/?iid=sr-link1

[13] Varol, O., Ferrara, E., Davis, C. A., Menczer, F., and Flammini, A. 2017. Online human-bot interactions: Detection, estimation, and characterization. In *Eleventh international AAAI conference on web and social media*.

[14] Clark, E. M., Williams, J. R., Jones, C. A., Galbraith, R. A., Danforth, C. M., and Dodds, P. S. 2016. Sifting robotic from organic text: a natural language approach for detecting automation on Twitter. *Journal of Computational Science*, 16, 1-7.

[15] Shu, K., Sliva, A., Wang, S., Tang, J. and Liu, H., 2017. Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter*, 19(1), pp.22-36.

[16] Pourghomi, P., Halimeh, A.A., Safieddine, F. and Masri, W., 2017, July. Right-click Authenticate adoption: The impact of authenticating social media postings on information quality. *In 2017 IEEE International Conference on Information and Digital Technologies (IDT)* (pp. 327-331).

[17] Safieddine, F., 2019, April. Facebook wants to combat fake news with ID checks – with 'grave implications' for our privacy. *The Conversation*. Available from: https://theconversation.com/facebook-wants-to-combat-fake-news-with-id-checks-with-grave-implications-for-our-privacy-116569 [Accessed 26/05/2019]

[18] Morley, C., 2008. Critical Reflection as a Research Methodology. *Knowing Differently: Arts-Based and Collaborative Research Methods,* pp.265-280.

[19] Alvesson, M. and Sköldberg, K., 2009. *Reflexive methodology: New vistas for qualitative research*. Sage.Vancouver.

[20] Fook, J., 2011. Developing critical reflection as a research method. In *Creative Spaces for Qualitative Researching* (pp. 55-64). SensePublishers.

[21] Fook, J., White, S. and Gardner, F., 2006. Critical reflection: a review of contemporary literature and understandings. *Critical reflection in health and social care*, pp.3-20.

[22] Osmond, J. and Darlington, Y., 2005. Reflective analysis: Techniques for facilitating reflection. *Australian social work,* 58(1), pp.3-14.

[23] Moon, J.A., 2013. *Reflection in learning and professional development: Theory and practice*. Routledge.

[24] Acquisti, A., Gross, R. and Stutzman, F., 2011. Faces of facebook: Privacy in the age of augmented reality. *BlackHat* USA, (2), pp.1-20.

[25] Earnshaw, R.A., de Silva, M. and Excell, P.S., 2019. Models of Research and the Dissemination of Research Results: the Influences of E-Science, Open Access and Social Networking. *Annals of Emerging Technologies in Computing (AETiC)*, 3(1).

[26] Dale, R., 2017. *NLP in a post-truth world.* Natural Language Engineering, 23(2), 319-324.

[27] Burkhardt, J. M., 2017. *Combating fake news in the digital age*. American Library Association.

[28] Haigh, M., Haigh, T., & Kozak, N. I., 2018. *Stopping fake news: The work practices of peer-to-peer counter propaganda*. Journalism Studies, 19(14), 2062-2087.

[29] Schultz, A., 2019. *How Does Facebook Measure Fake Accounts? Fakebook Newsroom*. [Online] [Accessed: 7th July 2019] Available from: https://newsroom.fb.com/news/2019/05/fake-accounts/

[30] P. Pourghomi, F. Safieddine, W. Masri, and M. Dordevic, (2017). How to stop the spread of misinformation on social media: Facebook plans vs. right-click authenticate approach. *IEEE Engineering & MIS (ICEMIS),* 2017, pp 1-8.

[31] Berghel, H., (2017). *Alt-News and Post-Truths in the" Fake News" Era*. Computer, 50(4), pp.110-114.

[32] Kreiss, D. and Mcgregor, S.C., 2019. The "Arbiters of What Our Voters See": Facebook and Google's Struggle with Policy, Process, and Enforcement around Political Advertising. *Political Communication*, pp.1-24.

[33] Navas, J. and Beltrán, M., 2019. Understanding and mitigating OpenID Connect threats. *Computers & Security,* 84, pp.1-16.