Risk Modeling of Disinformation;

Presented to Avril Haines & Alejandro Mayorkas

Executive Summary

Disinformation has and continues to exert a dangerous influence on civic life, national security, and public discourse. The incoming Biden administration must act to mitigate its harms while preventing further corrosion of public trust. Instead of engaging in an arms race of technology and competing information channels, we need to organize security personnel and institutions around threat detection while simultaneously stemming the spread of disinformation among our citizens. The United States' foreign intelligence and domestic security agencies under the new President should adopt a new strategy based on epidemiology. We can use Artificial Intelligence (AI) to build models of risk models while developing interventions for the people most likely to be exposed to and spread disinformation. Federal intelligence and security agencies should work directly with technology companies to build and deploy such a model to fight the battle on two fronts; first, by detecting sources of disinformation and, second, by mitigating the spread and (if possible) the effect of disinformation on people.

Background

Although there is debate about the effectiveness Russian disinformation campaigns to influence elections, we know that they are actively conspiring to subvert American democracy[1]. Foreign adversaries are very effective at weaponizing American companies like Facebook that use AI to target likely voters[2]. It's important to confront the Russian government about this espionage, but there might be little the U.S. can do to stop it. New threats are likely to emerge and proliferate following the successful incursion into the 2016 election. Simply naming and shaming, or persuing adversaries on their own terms will lead to a 'whack-a-mole' scenario draining the resources of Federal agencies. Moreover, technology and tactics will continue to advance as we remain reactive and defensive.

Warfare is rarely ever an open physical confrontation, but is frequently conducted through coordinated campaigns of sentiment engineering. The Internet Research Agency (IRA), the now infamous Russian troll farm, specializes in moral contagion[3]. Like other political advertising, politically motivated disinformation uses strongly presented moral emotional language that triggers people to engage and feel compelled to share. Russian espionage has perfected the art of outrage to push their agenda or simply to sow confusion and discord. If a

credulous Facebook user sees one of the ads featured in this New York Times article, they are more likely to engage with it compared to more cerebral or persuasive content. Social media gives everyone the potential for an unlimited audience. If a person feels morally compelled by a particular message, it would seem irresponsible not to share, opine, and request feedback from other users. To spread this important message, as it were.

Only a few memes ever become viral beyond the filter bubbles of their initial proponents. But the spectacular and enduring success of 'pizzagate' and conspiracy theories attached to the death of Seth Rich indicate to fabricators of such nonsense that the potential rewards are well worth the investment. It's important to keep in mind that sharing false information online is an exceptionally rare activity. Researchers at Princeton and New York University found evidence that suggests the propensity to disseminate 'fake news' can be narrowly attributed to a specific demographic profile[4]. Apart from the pernicious actors or foreign governments, most online disinformation is spread by people over 60 who lack the digital literacy necessary to discriminate truth from fiction. If we can anticipate and mitigate the vectors of transmission, then early intervention and contract tracing can slow the spread of viral disinformation.

Policy Recommendations

Research in AI and disinformation campaigns is advancing quickly, and it may be tempting to latch onto the latest trend or most startling piece of evidence revealed by the most recent investigation. At the same time, we cannot be reactive and waste energy debunking every false news story that gains traction with people who are unlikely to be persuaded or even to alter their sentiment in the face of provable deceit. To combat disinformation, we have to understand the motivations behind the people who spread it. Communication is fundamentally not about transmitting information, but transmitting values between parties[5]. People seek out like-minded compatriots who reflect their image of themselves and of the world they wish to inhabit. In high content saturation digital environments like social media, AI can be used against the interests of society by fulfilling the short term interests of individuals. We can deploy AI as part of a broader solution to protect against manipulation by malicious actors by adopting the following recommendations:

1) Create a body similar to the 9/11 Commission to fully investigate the full extent of foreign interference in U.S. elections. Until the full resources of the Federal government unhindered by a compromised Executive are deployed to this task, we cannot have a true public reckoning of the events leading up to the 2016 election.
2) Partner with social media companies to develop models of propensity and harm. Building on the lessons of the Covid-19 pandemic, we can test scenarios that act like stress tests of technology platforms. The digital and epidemiological meaning of the word 'viral' is informative here. By considering propensity indicators like those evaluated in the cited research above, the government in partnership with Facebook and Twitter could play out 'war game' scenarios of how disinformation may spread within a network comprised of high propensity individuals[6].
3) Hold social media companies accountable for their role in undermining democracy. Through active partnership, the government can exert more positive regulatory influence on digital platforms by compelling them to uphold their end of contracts and legal arrangements made with Federal agencies.
4) Implement strategic interventions to detect incursions and mitigate their harm. Once we have gained better knowledge following recommendation 1, proactively monitor attempts at political interference or social discord. Social media companies can reach out to individuals directly when they interact with disinformation and give them tools to better determine truth from fiction online.
5) Develop a rating system to label high quality trusted information sources.

Conclusion

The incoming Biden administration has an historic opportunity not just to repair the harms committed by the previous President and his foreign enablers but to proactively guard against future interference in American democracy. Implementing the recommendations above will restore confidence in U.S. elections and stem the tide of disinformation that poisons public discourse. We can use AI as a tool to model the risk of disinformation taking over the digital world, or we can remain defensive, hoping that another 2016 election happens while our enemies plot our collapse from within.

References:

1. "Grand Jury Indicts Thirteen Russian Individuals And Three Russian Companies For Scheme To Interfere In The United States Political System". 2018. Justice.Gov. https://www.justice.gov/opa/pr/grand-jury-indicts-thirteen-russian-individuals-and-three-russian-companies-scheme-interfere.

2. Tiku, Nitasha. 2017. "How Russia 'Pushed Our Buttons' With Fake Online Ads". Wired. https://www.wired.com/story/how-russia-pushed-our-buttons-with-fake-online-ads/.

3. Brady, William J., et al. "Emotion shapes the diffusion of moralized content in social networks". 2017. Vanbavellab.Hosting.Nyu.Edu. https://vanbavellab.hosting.nyu.edu/documents/Brady.etal.2017.PNAS.pdf.

4. "How Russia'S Troll Farm Is Changing Tactics Before The Fall Election". 2020. Nytimes.Com. https://www.nytimes.com/2020/03/29/technology/russia-troll-farm-election.html.

5. Guess, Andrew, Jonathan Nagler, and Joshua Tucker. 2019. "Less Than You Think: Prevalence And Predictors Of Fake News Dissemination On Facebook". Science Advances 5 (1): https://advances.sciencemag.org/content/5/1/eaau4586.

6. Rokhman, Lailiya. 2020. "James W. Carey - Communication As Culture Essays". Academia.Edu. https://www.academia.edu/7446970/James_W_Carey_Communication_as_Culture_Essays.

7. American, Scientific, Misinformation Disorder, and Claire Wardle. 2020. "Misinformation Has Created A New World Disorder". Scientific American. https://www.scientificamerican.com/article/misinformation-has-created-a-new-world-disorder/.

8. "Can The Michelin Model Fix Fake News?". 2017. The Daily Beast. https://www.thedailybeast.com/can-the-michelin-model-fix-fake-news.