STATEMENT OF
JOAN DONOVAN, PHD
DIRECTOR OF THE TECHNOLOGY AND SOCIAL CHANGE RESEARCH PROJECT AT HARVARD KENNEDY SCHOOL'S SHORENSTEIN CENTER ON MEDIA, POLITICS AND PUBLIC POLICY

HEARING ON "AMERICANS AT RISK: MANIPULATION AND DECEPTION IN THE DIGITAL AGE"

BEFORE THE SUBCOMMITTEE ON CONSUMER PROTECTION AND COMMERCE OF THE COMMITTEE ON ENERGY AND COMMERCE

DECEMBER 5, 2019

Online fraud is a great deal more widespread than many understand. Beyond malware, spam, and phishing attacks, beyond credit card scams and product knock-offs, there is a growing threat from new forms of identity fraud enabled by technological design. Platform companies are unable to manage this alone and Americans need governance.[1]

Online deception is now a multimillion-dollar global industry. My research team tracks dangerous individuals and groups who use social media to pose as political campaigns, social movements, news organizations, charities, brands, and average people. This emerging *economy of misinformation* is a threat to national security. Silicon Valley corporations are largely profiting from it, while key political and social institutions are struggling to win back the public's trust.[2]

Platforms have done more than just given users a voice online. They have effectively given them the equivalent of their own broadcast station, emboldening the most malicious among us.[3] To wreak havoc with a media manipulation campaign, all one bad actor needs is motivation. Money also helps. But that's enough to create chaos and divert significant resources from civil society,

[1] Klonick, Kate, "The New Governors: The People, Rules, and Processes Governing Online Speech." 131 *Harv. L. Rev*. 1598. https://harvardlawreview.org/wp-content/uploads/2018/04/1598-1670_Online.pdf

[2] Funke, Daniel, Susan Benkelman, and Cristina Tardáguila. 2019. "Factually: How Misinformation Makes Money." *American Press Institute*. https://www.americanpressinstitute.org/fact-checking-project/factually-newsletter/factually-how-misinformation-makes-money/.

Vaidhyanathan, Siva. 2018. *Antisocial Media: How Facebook Disconnects Us and Undermines Democracy*. New York, NY, United States of America: Oxford University Press.

[3] Glaser, April. 2019. "Bring Back the Golden Age of Broadcast Regulation. Especially for YouTube and Facebook." *Slate Magazine*. https://slate.com/technology/2019/06/youtube-facebook-hate-speech-regulation-how.html.

politicians, newsrooms, healthcare providers, and even law enforcement, who are tasked with repairing the damage.[4] *We currently do not know the true costs of misinformation.*

Individuals and groups can quickly weaponize social media to cause others financial and physical injury. For example,

1. Fraudsters using President Trump's image, name, logo, and voice have siphoned millions from his supporters by claiming to be part of his re-election coalition.[5] In an election year, disinformation and donation scams should be a concern for everyone.[6]

2. Along with my co-researchers, Brian Friedberg and Brandi Collins-Dexter, I have studied malicious groups, particularly white supremacists and foreign actors, who have used social media to inflame racial divisions.[7] Even as these imposters are quickly identified by the communities they target, it takes time for platforms to remove inciting content. [8] A single manipulation campaign can create an incredible strain on breaking news cycles, effectively turning many journalists into unpaid content moderators and drawing law enforcement towards false leads.[9]

Specific features of online communication technologies need regulatory guardrails to prevent them from being used for manipulative purposes. These include:

1. Registering, buying, and selling fake accounts, comments, and reviews to generate artificial attention, sometimes using botnets and automated text-generators to game algorithmic systems;[10]

[4] Bradshaw, Samantha, and. Howard, P. "The Global Disinformation Order: 2019 Global Inventory of Organised Social Media Manipulation." Working Paper 2019.3. *Oxford, UK: Project on Computational Propaganda*. https://comprop.oii.ox.ac.uk/research/cybertroops2019/

[5] Severns, Maggie. 2019. "Trump Campaign Plagued by Groups Raising Tens of Millions in His Name." *Politico*. https://www.politico.com/news/2019/12/23/trump-campaign-compete-against-groups-money-089454.

[6] Benkler, Yochai, Robert Faris, and Hal Roberts. 2018. *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*. Oxford University Press.

[7] Friedberg, B., & Donovan, J. 2019. On the Internet, Nobody Knows You're a Bot: Pseudoanonymous Influence Operations and Networked Social Movements. *Journal of Design and Science*, (6). https://doi.org/10.21428/7808da6b.45957184

Collins-Dexter, B. 2019. "The Dangers of Weaponized Truth." *Journal of Design and Science,* (6). https://jods.mitpress.mit.edu/pub/273294u8

[8] Donovan, Joan. 2019. "Opinion | First They Came for the Black Feminists." *The New York Times*. https:/www.nytimes.com/interactive/2019/08/15/opinion/gamergate-twitter.html

[9] Donovan, Joan. 2019. "How Hate Groups' Secret Sound System Works." *The Atlantic*. March 17, 2019. https://www.theatlantic.com/ideas/archive/2019/03/extremists-understand-what-tech-platforms-have-built/585136/.

[10] Caplan, Robyn, Lauren Hanson, and Joan Donovan. 2018. "Dead Reckoning: Navigating Content Moderation After 'Fake News.'" *Data & Society*. https://datasociety.net/output/dead-reckoning/

2. advertising products designed to inflate engagement metrics and/or force misinformation into users' search returns, feeds and timelines;[11]

3. networked factions (groups of loosely affiliated actors) strategically coordinating harassment, distributing hateful content, or inciting violence for profit or political ends;[12]

4. misusing platforms' donation features to raise funds for dangerous or imposter groups;[13]

5. promoting misinformation about health care to sell harmful or ineffective treatments; and[14]

6. using deceptively edited audio/video, like "deep fakes" and cheap fakes, to drive clicks, likes, and shares.[15]

Regarding the last point, the AI technology commonly called 'deep fakes' presents an immediate identity threat. Deep fakes are audio and video that realistically depict a person saying and doing things that never happened.[16] Social media companies are devising policies to prevent deep fakes

Confessore, Nicholas, Gabriel J. X. Dance, Rich Harris, and Mark Hansen. 2018. "The Follower Factory." *The New York Times*. https://www.nytimes.com/interactive/2018/01/27/technology/social-media-bots.html, https://www.nytimes.com/interactive/2018/01/27/technology/social-media-bots.html

Woolley, Samuel C. and Philip N. Howard. 2016. "Automation, Algorithms, and Politics| Political Communication, Computational Propaganda, and Autonomous Agents — Introduction." *International Journal of Communication* 10(0):9. https://ijoc.org/index.php/ijoc/article/view/6298

[11] Braun, Joshua A., and Jessica L. Eklund. 2019. "Fake News, Real Money: Ad Tech Platforms, Profit-Driven Hoaxes, and the Business of Journalism." *Digital Journalism* 7 (1): 1–21. https://doi.org/10.1080/21670811.2018.1556314.

Noble, Safiya Umoja. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: NYU Press.

[12] Donovan, Joan and Brian Friedberg. 2019. "Source Hacking: Media Manipulation in Practice." *Data & Society*. https://datasociety.net/output/source-hacking-media-manipulation-in-practice/

Lukito, Josephine, Jiyoun Suk, Yini Zhang, Larissa Doroshenko, Sang Jung Kim, Min-Hsin Su, Yiping Xia, Deen Freelon, and Chris Wells. 2019. "The Wolves in Sheep's Clothing: How Russia's Internet Research Agency Tweets Appeared in U.S. News as Vox Populi." *The International Journal of Press/Politics*, December, 1940161219895215. https://doi.org/10.1177/1940161219895215.

[13] Koh, Yoree. 2018. "Hate Speech on Live 'Super Chats' Tests YouTube." *Wall Street Journal*. https://www.wsj.com/articles/hate-speech-on-live-super-chats-tests-youtube-1541205849

[14] Zadrozny, Brandy. 2019. "These Are the Fake Health News That Went Viral in 2019." *NBC News*. https://www.nbcnews.com/news/us-news/social-media-hosted-lot-fake-health-news-year-here-s-n1107466.

[15] Paris, Britt, and Joan Donovan. 2019. "Deepfakes and Cheap Fakes." *Data & Society* (blog). 2019. https://datasociety.net/output/deepfakes-and-cheap-fakes/.

[16] Paris, Joan Donovan, Britt. 2019. "Deepfakes Are Troubling. But So Are the 'Cheapfakes' That Are Already Here." *Slate Magazine*. June 12, 2019. https://slate.com/technology/2019/06/drunk-pelosi-deepfakes-cheapfakes-artificial-intelligence-disinformation.html.

from misrepresenting public figures and average citizens, but this does not mean companies will adequately enforce these terms of service and address the damage done to society.

For example, in a recent report, researchers found 96% of deep fakes are pornography mostly targeting women.[17] This poses troubling questions about harassment and consent.[18] Mary Anne Franks and Danielle Citron have advocated for laws prohibiting non-consensual images because the potential for profit, exploitation, and extortion is high.[19] Unfortunately, even the most cutting-edge detection technology can be fooled by skillful deep fakes. For that reason, we need governance.

My co-researcher Britt Paris and I argue that so-called 'cheap fakes' are a wider threat. Like the doctored video of Representative Pelosi, last week's decontextualized video of Joe Biden seemingly endorsing a white supremacist talking-point poses a substantial challenge.[20] Because the Biden video was clipped from non-augmented footage, platforms refused to take down this cheap fake. Millions have now seen it. Platforms, like radio towers, provide amplification power and as such they have public interest obligations.

The world online is the real world, and *this crisis of counterfeits* threatens to disrupt the way Americans live our real lives. Right now, malicious actors jeopardize how we make informed decisions about who to vote for and what causes we support, while platform companies' own products facilitate this manipulation, placing our democracy and economy at significant risk.[21] What makes manipulated content so dangerous is the ease of distribution and the hidden protocols of moderation.[22]

[17] Ajder, Henry, Giorgio Patrini, Francesco Cavalli, and Laurence Cullen 2019. "The State of Deepfakes: Landscape, Threats, and Impact." *Deep Trace Labs*. https://deeptracelabs.com/mapping-the-deepfake-landscape/

[18] Chesney, Robert and Citron, Danielle Keats, "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security." 2019. *California Law Review* 1753. https://ssrn.com/abstract=3213954

[19] Danielle K. Citron & Mary Anne Franks. 2014. "Criminalizing Revenge Porn" 49 *Wake Forest Law Review* 345. https://scholarship.law.bu.edu/faculty_scholarship/643

[20] PBS. 2020. "How 2020 Candidates Are Grappling with Online Disinformation." *PBS NewsHour*.. https://www.pbs.org/newshour/show/how-2020-candidates-are-grappling-with-online-disinformation.

[21] Charlet, Katherine, and Citron, Danielle. 2019. "Campaigns Must Prepare for Deepfakes: This Is What Their Plan Should Look Like." *Carnegie Endowment for International Peace*. https://carnegieendowment.org/2019/09/05/campaigns-must-prepare-for-deepfakes-this-is-what-their-plan-should-look-like-pub-79792.

Acker, Amelia, and Donovan, Joan. 2019. "Data Craft: A Theory/Methods Package for Critical Internet Studies." *Information, Communication & Society* 22(11):1590–1609. https://doi.org/10.1080/1369118X.2019.1645194

[22] Gillespie, Tarleton. 2018. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. New Haven: Yale University Press.

Roberts, Sarah T. 2019. *Behind the Screen: Content Moderation in the Shadows of Social Media*. New Haven, CT: Yale University Press.

Roberts, Sarah. 2017. "Social Media's Silent Filter." The Atlantic. March 8, 2017. https://www.theatlantic.com/technology/archive/2017/03/commercial-content-moderation/518796/.

We must expand the public understanding of technology by guarding consumer rights against technological abuse, including a cross-sector effort to curb the distribution of harmful and manipulated content. As danah boyd and I have written, platform companies must address the power of amplification—separately from content— so that media distribution is transparent and accountable.[23] I urge Congress to do the same. Platforms have politics.[24] Regulation and technology must work in tandem, or else *the future is forgery*.

[23] Donovan, Joan, and boyd, danah. 2019. "Stop the Presses? Moving From Strategic Silence to Strategic Amplification in a Networked Media Ecosystem:" *American Behavioral Scientist*, September. https://doi.org/10.1177/0002764219878229

[24] Gillespie, Tarleton. 2010. "The Politics of 'Platforms'." *New Media & Society* 12 (3): 347–64. https://doi.org/10.1177/1461444809342738.