Media and Information Quality

Ethics and Governance of Artificial Intelligence Initiative

Social media platforms have long been celebrated as the core of our digital "marketplace of ideas," democratizing the creation and sharing of information—everything from cute cat pictures to breaking news. More recently, the same platforms have also been home to a darker side of the Internet: hate speech, online harassment and bullying, "fake news" and propaganda. These have become some of the most pressing information quality problems, and they show how vulnerable these algorithmically-governed sites are to manipulation and abuse. The scale and impact of these problems only underscore how far we are from fully understanding the depth and scale of these challenges, and how hard it is to develop solutions.

Social media platforms are designed to promote user engagement, hold attention, and maximize advertising revenue. Powered by secret algorithms, personalized newsfeeds are optimized to maximize clicks, with the unintended consequence that authentic commentary is mixed with with eye-catching false and misleading reporting. At a moment where two out of three Americans get news from social media, we do not know what proportion of this 'news' is intentionally false and misleading. Because each user sees unique, personalized content, it is nearly impossible to measure the proportions of content grounded in journalistic standards of objectivity and accuracy versus content designed to misinform the public and distort legitimate debate.

The 2016 U.S. election gave a taste of both the magnitude and importance of the issue. Inklings of Russia's interference trickled out in the weeks and months before election day. But it is only in last few months that revelations of coordinated bot armies and the influence of false news narratives have underscored the threat. Many questions remain about the role of AI in creating, propagating, and combatting media manipulation, but recent events demonstrate the vulnerability of our information ecosystem and our democratic systems. Because these vulnerabilities are as much human and social as they are technical, solutions will be found at the nexus of code and policy.

The information quality problem extends far beyond the promotion of deliberately fabricated reporting in the context of elections. The same automated systems provide an avenue for targeted attacks on individuals, harmful speech directed at groups based on race, gender, and ethnicity, and cynical campaigns to exacerbate social divisions for political ends. The scale and rapidity of online discourse in combination with vexing definitional issues—manipulative and harmful speech are hard to distinguish from protected speech—make these issues particularly challenging. **It is now clear that this is indeed a Frankenstein problem; social media platforms, regulators, and the public alike are struggling to fully understand the sources and depth of the problems and develop balanced, scalable solutions to manage these issues.**

BERKMAN KLEIN CENTER
FOR INTERNET & SOCIETY
AT HARVARD UNIVERSITY

23 Everett St., 2nd Floor
Cambridge, MA 02138

cyber.harvard.edu
hello@cyber.harvard.edu
@BKCHarvard
617.495.7547

In a world of over one billion daily Facebook users and hundreds of millions of YouTube, Twitter, WhatsApp, Instagram, and Snapchat users, harmful speech on these mediums continues to grow, highlighting the limitations of current ad hoc and patch-work approaches to mitigate its occurrence and deleterious effects. Meanwhile, AI will continue to shape the ways in which it propagates. What happens when someone is harassed by an orchestrated bot army, rather than a single person who can easily be blocked? Staying on top of such scenarios requires proactive and comprehensive approaches that combine deep understandings of policy and technology.

The **Berkman Klein Center** and **MIT Media Lab** are developing foundational tools, research, and communities to explore and address the impact of automation and machine learning on content production, dissemination, and consumption. This work is informed by their longstanding collaboration on Media Cloud, an open platform for the qualitative and quantitative study of online media ecosystems. Media Cloud ingests stories from over 25,000 media sources daily, rendering them available for analysis and mapping to elicit hidden relationships between topics, stories, media sources, and agendas in the data.

The Media and Information Quality track brings diverse stakeholders together to map the effects of automation and machine learning on content production, dissemination, and consumption patterns, while evaluating the impact of these technologies societal attitudes and behaviors and democratic institutions. We will develop tools that help both users and platforms better respond to these challenges.

Challenges that we seek to address in our work include:

› **Media Manipulation:** We will develop analytical tools and conduct empirical research to increase understanding of the mechanisms by which media manipulation occurs, track its prevalence within media ecosystems, and assess the potential for alternative interventions.

› **Harmful Speech:** We will focus on developing better research tools for tracking the propagation of harmful speech, and on improving the performance and application of tools being used to address harmful speech.

› **Platform Engagement:** Partnerships with a diverse group of stakeholders will inform research initiatives and potential policy interventions. Our work with platforms on issues such as information fiduciaries and forums such as the Partnership on AI will establish feedback mechanisms between social media companies and researchers.

Pillars of Impact

In building solutions that address these challenges, our institutions are focused on:

1. **Analytical Tool Development:** Including social network analysis, qualitative media analysis, and natural language processing to more effectively study these topics.

2. **Empirical Research & Case Studies:** Research efforts on better understand the spread of propaganda across different platforms and influence of automated tools on media manipulation and harmful speech online, including international contexts.

3. **Collaborative Problem Solving:** We continue to convene scholars from across disciplines, activated by the milestone Information Disorder, New Media Ecosystems and Democracy conference held in June 2018. Our approach bridges sectors and disciplines to enable development of the most robust techniques possible.

About the Ethics and Governance of Artificial Intelligence Initiative

The rapidly growing capabilities and increasing presence of AI-based systems in our lives raise pressing questions about the impact, governance, ethics, and accountability of these technologies around the world. How can we narrow the knowledge gap between AI "experts" and the variety of people who use, interact with, and are impacted by these technologies? How do we harness the potential of AI systems while ensuring that they do not exacerbate existing inequalities and biases, or even create new ones? At the Berkman Klein Center, a wide range of research projects – including the one outlined above – community members, programs, and perspectives seek to address the big questions related to the ethics and governance of AI under the Ethics and Governance of AI Initiative, launched in 2017.