Introduction

Critical attention required on the increasing sophistication and scope of disinformation

ERIC HORVITZ, CHIEF SCIENTIFIC OFFICER

Disinformation refers to the deliberate use of false information with the intention of influencing public opinion. Efforts to fabricate falsehoods for the purposes of manipulating the masses have a long history. However, new forms of disinformation have come to the fore over the last decade, enabled by advances in computing methods and infrastructure that have transformed the power, scope, and efficiency of disinformation campaigns.

Widely used consumer platforms and services, such as social media, creator platforms, search engines, and messaging services, now provide state and non-state actors with powerful channels for distributing disinformation. Beyond channels, these services provide malevolent actors with ready-made tools to experiment, monitor, iterate, and optimize the impact of disinformation campaigns.

Commercial online platforms have been harnessed by these actors as engines of disinformation to power messaging programs aimed at political influence, polarization, and chaos. Disinformation strategies are growing in sophistication, including the concerted use of multiple services[1] to reinforce messages across platforms.

On a second front, advances in machine learning (ML) and graphics have led to widely available tools for fabricating high-fidelity audiovisual content, referred to as synthetic media and deepfakes. For decades, photos and comments by political leaders have been manipulated or taken out of context in disinformation efforts, often with dramatic effects. However, technologies for generating deepfakes are providing malevolent actors with powerful, general palettes for fabricating behaviors and events. These methods are injecting new powers of persuasion into disinformation campaigns.

In a third area of concern, artificial intelligence (AI) methods can be used by state and non-state actors to formulate and drive powerful psychological operations that leverage insights and data about human cognition. ML and reasoning can be used to profile individuals and groups and to generate personalized programs of disinformation aimed at influencing beliefs, opinions, and actions.

The repurposing of consumer computing infrastructure, use of tools for generating synthetic media, and harnessing AI to guide psychological operations are troubling separately and in synergy. Together, they are supercharging disinformation, with grave implications for the health and vibrancy of democracies that depend critically on educated and aware citizenries.

What might we do in the face of these developments?

---

[1] A. Spangher, G. Ranade, B. Nushi, A. Fourney, E. Horvitz (2020). *Characterizing Search-Engine Traffic to Internet Research Agency Web Properties*, Web Conference.

We need to critically attend to the increasing sophistication and scope of disinformation—and to engage on multiple fronts. First and foremost, we need to invest deeply in modern media literacy, to educate people about how to understand, expect, and recognize disinformation and misinformation. Work on media literacy extends beyond education and includes efforts to provide new kinds of tools that can help people to critique the source and veracity of news and information. Second, we need to support high-quality journalism, including trusted news organizations. It is essential to have committed reporters on the ground to see, hear, and report with clarity on events and incidents. In addition, we need to assure the health and vibrancy of local journalism.

On the technical front, there is promise in applying AI pattern recognition technologies to detect patterns of communications and content that reveal an intent to deceive. Such work includes efforts to identify audiovisual and text-based media as fabricated. On another front, efforts in networking technologies can be aimed at identifying primary locations and organizational sources of disinformation. Finally, there are promising developments with technologies that employ a set of methods, including cryptography, security, and database technologies in production tools and pipelines that serve to certify the origin and history of edits to online media content, referred to as the *provenance*[2] of the content. Exciting progress with media provenance and authenticity is being nourished by strong cross-organization collaborations.

We are facing unprecedented disinformation campaigns and related cyber operations by state and non-state actors. These campaigns target public awareness and knowledge with disinformation, while others target enterprise operations and confidence. It is important to stay aware of developments and to come together to address the challenges with awareness, technologies, and policies. Addressing the new and evolving challenges will take ongoing investments, innovation, and activity on multiple fronts. This important chapter reviews some of these challenges and provides insights on directions forward

[2] A promising step forward on disinformation - Microsoft On the Issues