

# What's Worse Than Fake News? The Distortion Of Reality Itself.

**AVIV OVADYA** is chief technologist at the Center for Social Media Responsibility at the University of Michigan and a Knight News Innovation Fellow at the Tow Center for Digital Journalism at Columbia University.

Which hurts civilization more: no one believing anything, or everyone believing lies? If we fail to take immediate action to protect our news and information ecosystem, we may soon find out. We are careening toward an infocalypse—a catastrophic failure of the marketplace of ideas.

I am even more deeply concerned about the future of our democracy now than I was in mid-2016, when I was one of the few raising the alarm about social media creating an explosive breeding ground for misinformation. Facebook and its brethren have begun to take this threat seriously, but the next threat—the distortion of reality itself—is fast approaching.

Advances in communications technology and artificial intelligence are making it easy to create audio or video content with potentially dangerous consequences, from making it appear that a world leader is ordering a nuclear strike to simulating your spouse's voice on the phone asking for a bank password.

As individuals and institutions, we must ensure that we trust the trustworthy and disregard the disingenuous. Representative government requires accountability, and accountability requires discerning knowledge. But technology is disrupting that edifice of knowledge, and if we don't act quickly enough, it may soon bring us past the point of no return.

*We are careening toward an infocalypse—a catastrophic failure of the marketplace of ideas.*

*In the medical research and the drug industry, institutional review boards help evaluate new studies to ensure they are ethical. AI researchers and platform technologists need similar in-house or third-party societal impact review boards to help them evaluate the potential unintended consequences of their work.*

So what can we do? In short, we need massive investment across industry, civil society and government, to understand and mitigate threats to our information ecosystems. And we need it now.

The landscape of these threats will change rapidly over the next few years and will manifest in very different ways across different regions and systems. The funding vehicles, incentives and institutions that we develop to address them must therefore be adequately nimble and able to quickly reallocate resources as threats evolve.

As of now, there are a few particularly promising mitigations that deserve immediate consideration:

### **1. Monitoring the information ecosystem**

In order to react to emerging threats to our information ecosystem, we need systems in place to track and understand how it is changing, if it is improving and where it might need to be shored up. In 2016, for example, I was able use an off-the-shelf analytics tool to determine that a significant proportion of the most popular news consumed through online platforms was inaccurate or misleading.

With access to better data and investment in human judgment and fact-checking to establish baseline truths, this type of effort could be ongoing and respond to new challenges and platforms. This is the focus of my current work, as it informs all potential mitigations and can help us understand their effectiveness.

### **2. Fostering responsible research and design**

As machine learning technology advances and technology platforms scale, our information ecosystem is being deeply impacted. Organizations that are researching or deploying these new systems need to be able to understand the potential consequences of their creations. Unfortunately, there is no established body of experts with the motive and resources to evaluate how these systems could impact the world. In the medical research and the drug industry, institutional review boards help evaluate new studies to ensure they are ethical. AI researchers and platform technologists need similar in-house or third-party societal impact review boards to help them evaluate the potential unintended consequences of their work.

### 3. Implementing authenticity infrastructure

We need to proactively develop systems that can efficiently determine if, for example, a video is real or manipulated and convey this information to the end user. This is especially important for journalists, diplomats, courts and political leaders. Initially, forgery detection may work well enough.

Very soon, however, we may need to build and deploy an enormous amount of new technical infrastructure to proactively validate that a video was captured at a particular time and place. Either way, platforms and even browsers would need to incorporate these authenticity stamps in a way that has real impact for users. They must affect the psychology of belief formation, as merely labeling something as forged is not likely to be sufficient.

### 4. Ensuring information markets reward reality

Our current information ecosystem still strongly rewards misinformation, sensationalism and divisiveness over reality, accountability and civic connection. Whether you are a Russian troll, a Myanmar monk preaching ethnic cleansing or even an established newspaper, that's where the attention and money lies. We need to fix this at the distribution platform level and perhaps even at the societal level. This may require anything from changing how popularity of content on Facebook equates to more prominent play in news feeds to improving media literacy to creating a fund to reward those who catch forgeries.

These proposals are just the tip of the iceberg—and a dimly seen iceberg at that. Each is a huge endeavor, full of nuance and minefields, and will have many false starts that fail to bear fruit. But this is a battle we must fight. It's a future we must invest in if we want to avert an infopocalypse.

I know I would like to live in a functioning civilization. Wouldn't you?



*Very soon, however, we may need to build and deploy an enormous amount of new technical infrastructure to proactively validate that a video was captured at a particular time and place.*