RESPONDING TO FAKE NEWS THROUGH REGULATION AND AUTOMATION

Samantha Bradshaw
Oxford Internet Institute, Computational Propaganda Unit

Monetising clicks – incentivising junk news

One of the systemic factors underlying the burgeoning Fake News economy is that the technology and social media companies monetise our attention. If a content creator can deliver edgy information that attracts attention and encourages users to 'click through' to a story, they can generate advertising revenue. Fake News stories are often written with emotional appeal — to spread based on vitality opposed to veracity — and there are clear monetary incentives in place to maximise the production and distribution of this information. If social media companies incentivised high quality content — as opposed to emotionally appealing or outrageous stories and information — we may not have quite the same problem we're having now, where there is a race to the bottom in terms of content quality.

A swarm of bots and cyborgs

Recently, there has been heightened attention around how bots are distorting conversations on social media. In elections and referenda around the world, bots have been used to artificially drive up user engagement by liking, sharing, or retweeting content. Automating these interactions can serve to generate a false sense of popularity or consensus — not only around traditional consumer products, but also around political ideologies or individual beliefs.

Not all bots are bad and many remain an integral part of the internet ecosystem. Originally, bots were developed to perform repetitive and mundane tasks, such as conducting network maintenance or organising and cataloguing content. However, bot functionality was also extended to human interactions through internet Chat Relays, customer service tools, and social media interactions. Today, there are a variety of different terms and functions for these automated accounts, from harmless web crawlers to more malicious bots that are used to spread spam or disinformation.

Over the past two years, bots have been used to push polarizing messages to voters throughout the United States and Europe. Social media companies — such as Twitter, Facebook, and Google — have increasingly become concerned about the proliferation of bots on their platforms and have taken several steps to remove these accounts. There are a number of incentives to remove "bad" bots from online spaces as they not only undermine the quality of legitimate user interactions, but also the quality of user data that is sold to advertisers who want to reach real consumers.

One thing that many people do not realize is that there is an entire political economy supporting the buying and selling of botnets that can be put to purpose. These services are not only found in the deep corners of the internet's "Dark Web", but also on the mainstream internet where 1000 followers costs on average £20. As innovation continues in areas such as artificial intelligence and machine learning, bots will become increasingly sophisticated, making their detection and removal by platforms even more difficult.

Cont...

Planned offensives

Extremely effective disinformation campaigns involve careful planning, but in general there is a low barrier to entry. During the 2016 US election, disinformation stories were crafted and disseminated alongside traditional offensive cyber operations—such as email hacks and data leaks. For an attack this large, it would have taken months of preparatory work to identify networks of people and engineer situations to gain access to email accounts and sensitive documentation. It also takes time to execute a dissemination strategy, which typically involves strategic data leaks, crafting conspiracy theories, and the propagation and amplification of messages by bots until they are ultimately taken up elsewhere in the blogosphere, partisan media, and mainstream media.

'Pizzagate' is a clear example of an offensive that deployed all the elements discussed above. Email leaks — secured through phishing attacks — targeted Hillary Clinton's campaign manager John Podesta. In the raft of hacked emails were receipts from a pizza diner called Comet Pizza in Washington DC. These receipts eventually formed the basis of an online conspiracy that suggested John Podesta and Hillary Clinton were running a paedophile ring in the basement of this pizzeria. This conspiracy was amplified so broadly that a man named Edgar Welch drove to the pizzeria with a gun, fired shots in the air during business hours, and proceeded to search for children who were "trapped" in this basement. No one was hurt, and Welch was arrested and sentenced to four years in prison. Nevertheless, this example demonstrates the power of a well-executed disinformation campaign.

Clamping down: regulatory responses

Tackling Fake News is no easy task for government regulators. Increasingly, policymakers around the world are searching for new ways to deal with the spread of bad information online. However, new regulations that seek to thwart the spread of "false" or "fake" content could have a chilling effect on free speech. Instead, governments should look towards mechanisms that would encourage the enforcement of laws that are already in place to deal with harmful forms of content. Other initiatives could focus on the deeper systemic issues, such as how social media algorithms incentivize the spread of false, extreme, or other forms of negative content.

Another area of regulatory intervention could look at the surveillance economy and how data is used to target political messages to individuals online. Unscrupulous political and state actors have already exploited user data to target people with political messages and advertisements. These posts — sometimes called dark advertisements — are often tailored to an individual, so the message one person sees could be very different from another. We have seen political advertisements targeting minority communities during the 2016 election in the United States and in the United Kingdom to suppress voting and political participation. Ultimately, dark advertisements polarise voters, lower trust in institutions, and degrade the quality of our democracy.

Governments are taking a number of positive steps to improve transparency in political advertising. Some positive interventions include verifying the identity of people and organisations purchasing advertisements on social media, and allowing users to see who and why they are being targeted by messages. Other legislation requires social media companies to create a public archive of all advertisements bought and sold, to hold political parties accountable for any dark advertisements they are purchasing during their campaigns.

Clamping down: computational responses

Artificial intelligence is often proposed as a solution to Fake News. There are a number of areas where AI is really effective in flagging, blocking, and removing content online. In areas such as child protection or terrorism, great strides have been made in applying AI and machine learning models to tackle the spread of harmful content. However, it is difficult to automate a response to "Fake News" because what is considered "truth" can be subjectively different for everyone, as opposed to issues related to child protection or terrorism where the decision to remove content is much more black and white. At the same time, most "Fake News" and propaganda is switching from simple text to video and images, where there is still limited AI capacity.

"...platform companies like Facebook and YouTube are concerned about the proliferation of fake accounts and bots, because it undermines the quality of the user data."

Nevertheless, there are indicators that could be used to flag different types of content and potentially to identify Fake News. For example, algorithms can down-rank content that individuals share but don't actually click through to read. Nevertheless, having AI make decisions about what content is true and what is false would be morally and politically perilous, resulting in a range of free speech issues that would de-value social media. Instead, computational responses are best suited to identifying instances of harmful, fake or conspiratorial content going viral and flagging them for action, but review by human editors should always remain a part of the take-down process.

Watching, learning, applying our best laws

Social media hasn't broken our legal and regulatory systems. There are still many laws that can be applied to protect individuals from hate speech, harassment, defamation, and other forms of harmful content. What is needed is better enforcement of existing legal structures, as well as the terms of service agreements developed by companies. This won't solve all problems around Fake News, but it will ensure that the internet remains a free and open space for ideas and conversations.

With every new technology there has always been a period of learning, and old laws now need to be updated for present times. The invention of the printing press, radio, and television all had similar learning periods where society updated its norms, regulations and laws to limit bad behaviour while reinforcing the good. With social media, we are currently in this learning phase.

New technologies always bring uncertainty, and it's not always immediately clear how our old laws can be updated to address some of the changes we're facing today. But laws are designed to be durable, and it's important that government, citizens, lawyers and industry have an open and active conversation about how law can mitigate harms while enhancing democracy.

13%

The number of Twitter accounts that are actually automated.