

## **AI-Generated Fake News and the Dilemma of Misinformation**

In the ever-expanding world of the internet and social media, perhaps one of the things that we encounter most often is false information - more specifically, fake news, false information, or over-exaggerated points from media outlets which are intended to evoke emotion instead of informing the reader. This type of content can be found whether the information is spread maliciously or unintentionally due to the easily-accessible nature of posting on social media sites. Regardless of the intention, and due to both the sheer volume of content and amount of unreliable news stories, disinformation can spread incredibly rapidly and gain traction and following as a result of the ease of sharing. This presents a serious problem when combined with the potential of AI-generated content and news.

Because of the way that human speech patterns are incredibly repetitive and quantifiable, specifically in writing, AI is able to mimic human writing patterns to a tee, and produce information as a result of language combinatory logic and massive amounts of data. A case study which was presented in New York Times illustrates a fantastic case study of the consequences of this - a company known as OpenAI has created an AI which generates news articles based on data that it is fed, mimicking English writing patterns and utilizing nonfiction events to generate artificial news content which is eerily similar to human writing. In fact, the content is so well produced that the authors of the article actually point out having trouble telling the real human writing from the AI writing in some cases. The authors show how when given a prompt to write about US political parties, for example, the content of the article dynamically changes depending upon the commonly held viewpoints of each party [1]. As it stands - these types of advanced content-generation programs are exclusively reserved for researchers and the companies that developed them, and are not being released to the general public.

Why does choice to release or not release this technology present an ethical dilemma? The authors of this article first point out that this presents a dangerous way for misinformation to be spread, especially on a wide scale. When this is in the hands of researchers or academics, it is less of an issue and more of an area of study about content generation, but if such a technology becomes open-source and available to the general public, this is where the problem begins to unfold. The NYT article points out if these tools are utilized by bad actors, content-generation en masse can actually be a powerful tool for spreading propaganda and disinformation.

On the other hand, others have argued that such content generation isn't inherently flawed. An article from The Verge points out that when moderated, such content being made available to the public can be a powerful tool for researchers [2], and the fear that it could be used negatively can perhaps be overstated due to current limitations with the technology. For one thing, they point out that OpenAI's generated articles aren't flawless, and moreover, can never replace previously credibly-established news organizations. Moreover, for uses like entertainment media, this can actually have huge advantages, as no-name sites and companies

Jake Henson

11/05/2020

CSCI 3202 Intro to AI

seeking to make traffic-driven ad-revenue are able to profit for actually very little work when utilizing similar algorithms to generate articles regardless of content, as one CNBC author notes [4].

One of the ways to examine this dilemma on a more abstract level is outlined in an UCS article on ethical studies [3]. Through the use of an ethical framework known as utilitarianism, the idea that ‘utility’ or overall ‘goodness’ ought to be maximized, reducing the amount of harm in the world. In the case, the question becomes - are the research opportunities and potential gains from releasing this technology worth releasing AI programs which can generate this type of content rapidly? An issue that can be raised with this is that the amount of overall harm done via the proliferation of fake articles and AI generated content being shared and spread by social media may, in fact, far outweigh the simple research opportunities and value gained from releasing such technology as open-source and giving access to the public. However, dissenters may argue that begs the question whether such AI generated fake-news could hypothetically be prolific enough to spread disinformation, i.e., would the net harm done actually be as prolific as some researchers are concerned about? If the technology isn’t perfect, and is only used within the sphere of less-consequential entertainment media for example, why worry? For this, one can argue, when combined with something like a botnet, for example, it can be used by people en masse to ostensibly drown out real information and news stories reaching social media before other companies report on issues. It can cloud the facts of a story with similar-sounding or similarly written but not objectively true content. This presents the issue of decreasing overall net-utility as a result of the technology being released.

When arguing the question of rights, whether people have a right to this fake-news AI, perhaps this is less suited for the scope of this paper, suffice to say that the question whether people have a right to AI driven-content is perhaps hard to fully answer without asking questions about the ownership of scientific discoveries and intellectual property.

The “common good” lens can also be used here, such that, is there a common-good for the public to be able to generate AI fake news? It is difficult to say, especially considering the fact that corporations can absolutely monetize this with traffic-based revenue, while the overall common good from simply reading this AI-created news may not actually be worth it. Of course, this all depends on the accuracy, intention, and methods by which this AI is used. AI generated news, if used as a tool for research, could actually benefit those seeking to combat fake news and identify methods by which AI-driven content could be classified, increasing the common-good to most peoples’ lives, such that they could ironically see less fake news.

This, in turn, leads us to view this dilemma through the lens of virtue [3] - when presented with the opportunity for massive amounts of AI driven content, will this be used for morally virtuous pursuits? Perhaps, one can argue, the proliferation of AI driven content with fake news can be used to *drown out* propaganda, or even used to quickly produce legitimate information when combined with factually-verified data to generate content. Again, citing the example previously, this technology could be utilized by researchers to combat fake news. On the other hand, we must ask ourselves whether such pursuits will, in fact, be followed by the

Jake Henson

11/05/2020

CSCI 3202 Intro to AI

general public and corporations valuing a profit, further abstracting the problem to whether people will act virtuously when presented with such powerful tools of content generation.

Regardless of the ethical framework though, there are several consequences which fall from both the action of continuing to generate such fake news and release the methods to do so to the general public. Even if researchers were able to identify the fake news now, with foreseeable improvements of the technology in the future, that line begins to blur. What then are they to do? Furthermore, when this technology is combined with botnets, massive media conglomerates, and the ever-expanding world of social media, it raises concerns about the legitimacy of information and being able to quickly verify and fact-check this news, which is already a growing problem.

To conclude, this type of technology can be perhaps ethically positive, in the sense that it can be monetizing non-essential news media like entertainment and combating fake news, however the potential for misinformation of truly important stories, political news, or bad actors to take over the content, or to simply flood the market with misinformation, regardless of intent, is simply too great. It doesn't benefit the common good, nor does it guarantee people will act in a virtuous way if this AI is able to be used by everyone. In an already over-saturated world of media content spread via social media, we need to ask ourselves whether utilizing AI to generate fake news articles is the most ethical thing to do. With the sheer volume of content able to be produced at any given moment means that given a proper outlet, such articles can spread to an absolutely absurd amount of people over such an amount of time, creating a veritable ocean of content which is impossible to keep up with and fact-check every detail.

As it stands, I would wholeheartedly fall within the argument that releasing this AI-content-generation tool to the general public is dangerous, due to the amount of potential utilitarian harm, the lack of common-good, and the questionable virtuousness of individuals who have this technology. The technology is still too much in its infancy phase, and combined with social media and the potentiality for bad actors to use this, presents a dangerous future where we can no longer tell the difference between real and fake news.

Jake Henson  
11/05/2020  
CSCI 3202 Intro to AI

## REFERENCES:

- [1] Cade Metz and Scott Blemethal, 2019. *How A.I. Could Be Weaponized to Spread Disinformation*. The New York Times. Retrieved from <https://www.nytimes.com/interactive/2019/06/07/technology/ai-text-disinformation.html>
- [2] James Vincent Feb 21, 2019. *AI researchers debate the ethics of sharing potentially harmful programs*. The Verge. Retrieved from Retrieved from <https://www.theverge.com/2019/2/21/18234500/ai-ethics-debate-researchers-harmful-programs-openai>
- [3] SCU. Santa Clara University, Markkula Center for Applied Ethics. *A Framework for Ethical Decision Making*. Retrieved from <https://www.scu.edu/ethics/ethics-resources/ethical-decision-making/a-framework-for-ethical-decision-making/>
- [4] Megan Graham, 2020. *To Show How Easy It Is For Plagiarized News Sites To Get Ad Revenue, I Made My Own*. CNBC, retrieved from <https://www.cnbc.com/2020/05/17/broken-internet-ad-system-makes-it-easy-to-earn-money-with-plagiarism.html>