

Opinion

Why We Need to Bust Some Myths about AI

Daniel Leufer^{1,2,3,*}

¹Access Now, Brussels Office, Belgium

²Mozilla Foundation, Mountain View, CA, USA

³Working Group on Philosophy of Technology at KU Leuven, Leuven, Belgium

*Correspondence: daniel.leufer@accessnow.org

<https://doi.org/10.1016/j.patter.2020.100124>

In this piece, Daniel Leufer introduces his project, aimyths.org, a website that tackles eight of the most harmful myths and misconceptions about artificial intelligence. By showing the damage caused by everything from misleading headlines about AI progress to attempts to use machine learning to predict complex social outcomes, Leufer's project underlines the importance of getting beyond the hype around AI.

Introduction

If, like me, you follow the latest news on technology, data science, and other related topics, chances are the you encounter one or more of the following at least once a day:

- misguided headlines that claim that artificial intelligence can accomplish some incredible task;
- sensationalist articles that misrepresent the results of scientific articles about machine learning benchmarks; and
- hyped-up claims about artificial intelligence (AI) being a silver bullet to solve complex social problems.

Each of these phenomena is the result of the hype, myths, and misconceptions that prevail about the discipline of AI. Some level of misconception about AI is understandable, given that precisely defining the aims of AI as a discipline is something which practitioners cannot even agree about. However, the majority of the misconceptions are harmful and avoidable, and some of them are arguably deliberately perpetuated by those who benefit from AI hype.

In March 2019, as I was beset by these infuriating phenomena, I came across the call for proposals for the Mozilla Fellowship. I ended up applying for a fellowship on the Open Web track, which is a collaboration between Mozilla and the Ford Foundation. Successful applicants can embed at a host organization to work on projects that promote Internet health. For 2019, the specific theme of the call for proposals was for projects to help achieve trustworthy AI. I decided to

make my frustration with AI hype into something productive and applied with a proposal for a project to investigate and tackle the worst misconceptions about AI.

My application was successful, and I was given the chance to embed at Access Now, a global digital rights NGO that fights to defend and extend the rights of users at risk around the world. Along with working to combat Internet shutdowns around the world and keeping individuals and organizations around the world safe online with its Digital Security Helpline, Access Now has been a leading voice in ensuring that the development of machine learning/AI is done in a way that protects our human rights.

Access Now also organizes RightsCon, the largest convening in the world of digital rights. At RightsCon 2019 in Tunis, I participated in a session that brought together a variety of stakeholders who work on AI policy: representatives from civil society, international institutions, governments, and companies. What connected everyone in the room was that we were all working to ensure that AI development and deployment respects human rights.

We also found that we all had a shared frustration: the huge amount of time we spent refuting misconceptions about AI, including attempting to clarify the vague term “artificial intelligence” and interrogating the idea that any regulation of AI will necessarily kill innovation.

We came out of that session with a list of some of the worst and most harmful misconceptions, and a clear sense that something had to be done to combat this wave of hype and misinformation. Fortunately, I received word soon after

that I had been granted a Mozilla fellowship, and I centered my project on developing resources to tackle these AI myths.

The Project

From October 2019 to July 2020, I set to work researching which myths to focus on and how best to tackle them. Following a blog post and survey to launch my project, it became clear that there was pretty general agreement on which myths got on people's nerves the most. From the survey results, there was a clear top 5:

- (1) AI has agency;
- (2) superintelligence is coming soon;
- (3) the term AI has a clear meaning;
- (4) AI is objective/unbiased; and
- (5) AI can solve any problem.

With a bit of editorial discretion to pick the 6th, 7th, and 8th myths (*ethics guidelines will save us, we can't/shouldn't regulate AI, and AI = shiny humanoid robots*, respectively), I now had a list of issues to tackle.

It was also clear to me that the resources to tackle these myths didn't have to be created from scratch: they already existed thanks to the great work of a host of researchers and activists, it's just that they weren't always in the most digestible format. My aim was to channel existing research, often from academic papers, into a more readable and adaptable format that could be useful for people who encounter this hype daily in their work.

Working with various experts, as well as an incredible developer and a fantastic designer, I put together aimyths.org, a site which tackles the eight myths



Headline rephraser

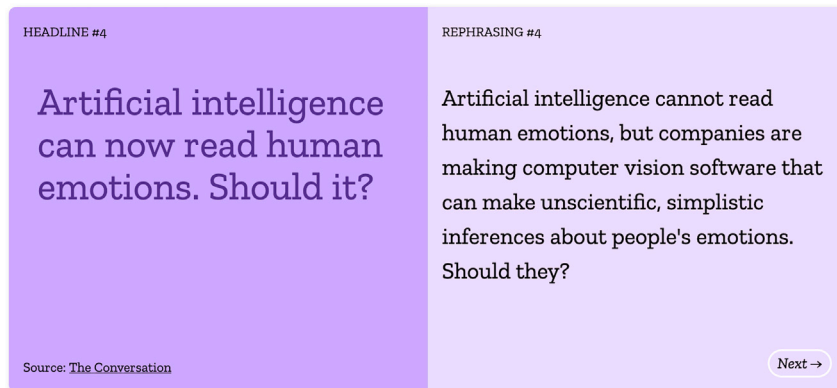


Figure 1. Headline Rephraser Tool from aimyths.org

mentioned above. One key idea that underlies the website is that none of these “myths” are straightforwardly false. The website tackles the prominent, superficial misunderstandings around AI that fall under these eight categories, but dealing with these issues doesn’t lead to straightforward refutations because there tends to be layers of complexity to each issue.

As an example, while it’s clear that claims from headlines about “AI developing its own language” are overblown, the problem of human agency versus machine agency is a tricky philosophical issue. However, each myth/misconception has its own particularities, so let’s look at a few of them in more detail.

AI Has Agency

The myth that got the most votes in the survey was that AI has agency. Now, this is a complex topic, which can quickly lead into philosophically and technically murky waters. However, there is one very prominent, harmful aspect of the issue that we are all familiar with: bad headlines. To take only the most recent controversial example, on September 8, 2020, the *Guardian* published an opinion piece entitled: “A robot wrote this entire article. Are you scared yet, human?”¹

The piece was attributed to GPT-3, the latest version of OpenAI’s impressive natural language generation tools. What the headline misrepresents, however, is the heavy-handed editorial work that went into the piece. In fact, the *Guardian* provided GPT-3 with a prompt containing the main elements of the article and got

the tool to produce eight different texts. The editors then took the best bits from these 8 and re-worked them into a single text (an update was published on September 11 clarifying this process²).

These essential details were only provided at the end in an editor’s note which most people are unlikely to read. The end result is that the general readership goes away with a grossly inflated perception of what “AI can do” that is impossible to live up to.

In such cases, the problem is that the ascription of agency to AI *masks the human agency behind certain processes*. This might seem relatively harmless in some cases, but most of the time we’re done a disservice when we’re presented with misleading claims about AI doing something when it is very clearly a case of humans using AI to do things. This is especially true in cases where machine learning systems are being used to make sensitive decisions about credit scores or social welfare payouts.

No AI system, no matter how complex or “deep” its architecture may be, pulls its predictions and outputs out of thin air. All AI systems are designed by humans—they are programmed and calibrated to achieve certain results, and the outputs they provide are therefore the result of multiple human decisions.

When we fail to see these decisions (or when they are deliberately masked), we look at AI systems as finished products, as systems that simply take input data and output objective results. This contributes to what Deborah G. Johnson and

Mario Verdicchio call *sociotechnical blindness* in their article *Reframing AI Discourse*, “What we call sociotechnical blindness, i.e., blindness to all of the human actors involved and all of the decisions necessary to make AI systems, allows AI researchers to believe that AI systems got to be the way they are without human intervention.”³

To help counter the socio-technical blindness that these headlines create, we developed a headline rephraser tool for aimyths.org which displays the original, misleading headline on one side, and then “debunks” it by rephrasing it more accurately (if somewhat cheekily) (Figure 1).

Terrible and Inappropriate Robots

A related, and equally prevalent, phenomenon to the bad AI headlines is the prominence of inappropriate and/or terrible pictures of humanoid robots that accompany news articles, governmental reports, and all types of other media about AI (Figure 2). I explicitly say inappropriate and/or terrible because we are, in fact, dealing with two separate (albeit, usually intersecting) phenomena.

The typical example of an inappropriate robot picture consists of a news article about some topic with an artificial intelligence component to it, but nothing to do with robotics, being illustrated by a picture of a (usually) humanoid robot (think of an article about AI in insurance being illustrated with a picture of a NAO robot). Such image-text combinations are in themselves harmless, and often quite hilarious for how misguided they are.

Finding “pure” examples of this category of “inappropriate placement of okay robot picture” is quite difficult, however, because in addition to being inappropriately placed, pictures of robots tend to be really, really terrible. In the myth *AI = shiny humanoid robots*, I break down the various ways in which robot pictures are terrible, offensive, and harmful.

One of the key ways that these pictures go wrong is through absurd sexualization. The robots in these stock photos often have inexplicable breasts and perpetuate harmful gender stereotypes by portraying “female” robots in sexualized or subservient roles. However, this sexualization of pictures of robots (and real-life robots) is just part of a huge problem of gender-based discrimination in the broader field

of AI. Much has been written about the harms caused by the gendering of AI assistants, as noted in UNESCO's report *I'd blush if I could*,⁴ which examines the harms caused by the gendering of AI voice assistants, and we have also seen how AI systems can perpetuate gender bias in hiring and allow for new forms of harassment against women.

Another problem with these images is that they are overwhelmingly white. As Dr. Beth Singler discovered in her investigation of the "AI Creation Meme" (the classic robot hand touching/shaking human hand⁵), in the 79 examples of this meme that she found, the human hand in all 79 cases was white, and, to come back to our previous point about gender, the human hand was male in 78 out of 79 cases. In their paper, *The Whiteness of AI*, Stephen Cave and Kanta Dihal note the seriousness of this overwhelming whiteness when they say that "AI racialized as White allows for a full erasure of people of color from the White utopian imagery."⁶ By centering whiteness as the default color of the future, these images contribute to envisioning a technological future that excludes people of color in much the same way that Big Tech today does.

Conclusion: Debunking Myths Is Just the Starting Point

Beyond the two myths discussed here that focus on the representation of AI, aimyths.org delves into policy issues around ethics and regulation of AI, as well as more technical and political issues



Figure 2. Typical Dodgy Robot Picture Used to Illustrate Articles about AI

such as understanding AI bias and the limits of machine learning approaches. While the website hopefully goes some way in debunking the most obvious and harmful misconceptions about AI, it by no means completes the work.

Debunking myths alone will not ensure that AI technologies are developed in a way that really serves people without undermining their rights and freedoms. Every week brings news of AI systems rolled out without consideration of the risks for the human rights of those impacted by them and without even the most minimal safeguards to prevent or mitigate harm. Worse, in many cases, from spurious gender detection apps to highly dangerous systems that claim to "predict criminality," those implementing the systems in question do not appear to consider whether the problems these "solutions" purport to address are even solvable using machine learning.

While critics can make every effort to counter hype and misconceptions, the responsibility ultimately rests with AI developers to develop and promote responsible, hype-free alternatives that would be worthy of people's trust and that would empower people instead of trying to bamboozle them with inflated claims.

REFERENCES

1. GPT-3 (2020). A robot wrote this entire article. Are you scared yet, human?, *The Guardian*, September 8, 2020 <https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3>.
2. Guardian US Opinion Editors (2020). How to edit writing by a robot: a step-by-step guide, *The Guardian*, September 11, 2020 <https://www.theguardian.com/technology/commentisfree/2020/sep/11/artificial-intelligence-robot-writing-gpt-3>.
3. Verdicchio, M. (2017). Reframing AI Discourse. *Minds Mach.* 27, 575–590.
4. UNESCO (2019). UNESCO: Explore the Gendering of AI Voice Assistants. <https://en.unesco.org/EQUALS/voice-assistants>.
5. Singler, B. (2020). The AI Creation Meme: A Case Study of the New Visibility of Religion in Artificial Intelligence Discourse. *Religions* 11, 253.
6. Cave, S., and Dihal, K. (2020). The Whiteness of AI. *Philos. Technol.* <https://doi.org/10.1007/s13347-020-00415-6>.

About the Author

Daniel Leufer works as Europe Policy Analyst at Access Now on issues related to artificial intelligence and data protection, with a focus on biometric surveillance. From October 2019 to July 2020, he was hosted by Access Now as a Mozilla Fellow, where he developed aimyths.org, a website that gathers resources to tackle eight of the most common myths and misconceptions about AI. He has a PhD in Philosophy from KU Leuven in Belgium and has worked on political philosophy and philosophy of technology. He is also a member of the Working Group on Philosophy of Technology at KU Leuven.