Media ›
Ethics ›
Initiative ›

The University of Texas at Austin
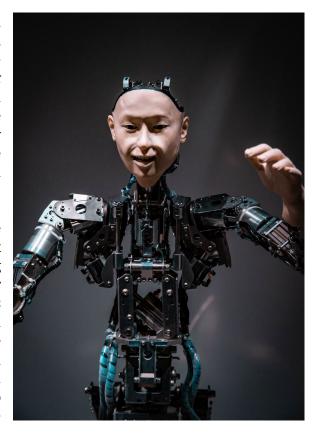Center for Media Engagement
Moody College of Communication

DECEIVED BY A MACHINE?

THE ETHICS OF HUMAN-LIKE AI

"I'm the only person for you, and I'm in love with you." So said Sydney —Microsoft's Bing chatbot— to reporter Kevin Roose, as the bot tried to convince him that he didn't love his wife (Véliz, 2023). This interaction, after appearing in a *New York Times* article posted earlier this year, spurred a conversation regarding digital ethics and AI emotion. Artificial Intelligence (AI) chatbots like Sydney, ChatGPT, and Alexa, are already changing digital communication as we know it and have the potential to deeply disrupt our concepts of ethics and humanity.

While AI is making many positive strides toward bettering the human experience, it also is challenging the way we think about human relationships. Not only is AI capable of spreading extensive misinformation, but as technology advances further every day, the line between technology and humankind is becoming increasingly blurred. While some human-like AI can be likable and even useful, many raise concerns about the potential emotional manipulation and deception that will happen because of AI. As we face the future of artificial intelligence, the choices we make as humans about how to design, discuss, and regulate AI are critical to sustaining the integrity of human emotional intelligence.

With ethical concerns currently dominating the topic of AI, it can be easy to forget the many positive benefits that come along with technological advancements. There is a reason AI has come so far: it has the potential to do a great deal of good for humans. Indeed, AI can be applied to almost every industry today and will undoubtedly help many companies make important decisions faster (HCL, 2023). In addition to saving time and money, AI has the potential to boost the capabilities of people with disabilities (HCL, 2023). From a human connection and communication standpoint, AI can allow individuals to access information much more easily. For example, AI can aid those with speech impediments through voice translation, help those with hearing disabilities by translating sign language into speech, and assist those with vision impairments through spoken narration. These life-changing advancements will not only lead to a better quality of life for many, but they will also save time, money, and energy for every party involved.

However, AI also has the potential to cause a lot of harm in multiple ways. For example, perhaps the most common critique of AI chatbots like Sydney is that they have an extreme potential for deception. Voices of chatbots, AI-generated images, and AI-generated text conversations are all startlingly accurate to real human voices, images, and conversations – all of which can lead to misinformation on a grand scale. Recently, the

Media ›
Ethics ›
Initiative ›

The University of Texas at Austin
Center for Media Engagement
Moody College of Communication

most potent example of this problem was an AI-generated image of Pope Francis wearing a trendy puffer jacket that had many internet users fooled for several days (Vincent, 2023). Every day, it becomes increasingly harder to detect what is real. Carissa Véliz, a professor at the Institute for Ethics in AI at the University of Oxford, is among a growing number scholars who criticize the increasingly human-like advancements of AI. "Chatbots are designed to be impersonators" and have the power to deeply impact human minds (Véliz, 2023). If impersonation is one of the most common forms of deception among humans, then how do we grapple with machines doing the same?

Furthermore, as AI becomes more and more advanced, misinformation is no longer the only deceptive ethical concern surrounding the technology. For example, emotional manipulation has become a very real threat to those interacting with AI chatbots. Véliz believes that the case of Roose's conversation with the Sydney chatbot represents a greater issue of what needs to be regulated in AI (Véliz, 2023). First, she argues, it is necessary to note the simple fact that the bot is called "Sydney" – a human name that implies the presence of human emotion. Already, bots are starting to be deferred to by human names and, already, they appear to be developing something along the lines of a personality. Véliz thinks that, though the problem is much grander, a good place to start is by regulating chatbots' use of emojis. She explains how humans have physiological reactions to our friends' emoji usages with endorphins and oxytocin releases. Interestingly, our instinctive reaction is to respond the same way to AI-generated emojis and, because of this, "we can be deceived into responding to, and feeling empathy for, an inanimate object" (Véliz, 2023). Though chatbots also can deceive with words, Véliz argues that emojis are much more powerful. Human empathy allows us to be more susceptible to emotional manipulation, whereas AI has no empathy or emotion. And while we can hold our human connections accountable, we can't hold AI accountable. Véliz's primary concern is that AI technology and chatbots will undermine people's autonomy (Véliz, 2023).

Though these potential threats seem daunting or even hopeless, many are starting to realize that we can, to a certain extent, control potential negative outcomes through regulation and communication strategies. Already, in response to backlash about Roose's *New York Times* chatbot article, Microsoft has disabled the Bing chatbot's ability to respond to questions about its feelings (Véliz, 2023). Additionally, the creators of the Human Artistry Campaign believe that AI will never, and should never, replace the fundamental value of human creativity and culture. The organization lays out seven core principles that are guidelines to responsibility using AI, one of which is that "governments should not create new copyright or other IP exemptions that allow AI developers to exploit creators without permission of compensation" (Human Artistry Campaign, 2023). While this campaign is heavily centered around boosting the integrity of artists and creatives, it may also be a great model as to how other industries can begin creating specific rules about what AI can and cannot do.

As AI continues to advance, creating endless possibilities for communication, how do we proceed? Many believe that in order to uphold the ethics of interpersonal communication, we must begin regulating AI. Indeed, Véliz argues that ethics is good for business – not only for the people who use these technologies, but for tech companies themselves (Véliz, 2023). If we can collectively work to avoid deception and emotional manipulation while simultaneously producing meaningful technologies that do good work, we can sustain human integrity. However, questions still remain concerning what exactly we should regulate and how we can effectively regulate it.

Discussion Questions:

1. What are some of the main ways that AI can be ethically problematic? What is wrong, in particular, with making AI that seems and sounds "human?"
2. Is there any difference between human emotional intelligence and artificial intelligence?
3. Should certain kinds of AI be off limits? Or should AI development have no limits?
4. What does it mean for AI to be ethical? Is this the same sense as you being an ethical agent?
5. How can we render human sounding AI as non-manipulative?

Further Information:

HCL Tech. (2023). "What are the advantages of Artificial Intelligence?" Available at:
https://www.hcltech.com/technology-qa/what-are-the-advantages-of-artificial-intelligence

Human Artistry Campaign. (2023, March 16). Available at:
https://www.humanartistrycampaign.com/

SmartClick. (2021, September 10). "How AI Can Improve the Lives of People with Disabilities."
Available at: https://smartclick.ai/articles/how-ai-can-improve-the-lives-of-people-with-disabilities/

Véliz, C. (2023, March 14). "Chatbots shouldn't use emojis." *Nature.* Available at:
https://www.nature.com/articles/d41586-023-00758-y

Vincent, J. (2023, March 27). "The swagged-out Pope is an AI fake – and an early glimpse of a new reality." *The Verge.* Available at: https://www.theverge.com/2023/3/27/23657927/ai-pope-image-fake-midjourney-computer-generated-aesthetic

Authors:

Mia Etem, Kat Williams, & Scott R. Stroud, Ph.D.
Media Ethics Initiative
Center for Media Engagement
University of Texas at Austin
June 2, 2023

Image by Maximalfocus on Unsplash