

ChatGPT-4 reinforces sexist stereotypes by stating a girl cannot “handle technicalities and numbers” in engineering

Leading AI policy and human rights experts invite OpenAI's Chief Exec Sam Altman to collaborate on tackling governance issues and dangers accompanying ChatGPT

NEW YORK, NEW YORK, USA, March 23, 2023 /EINPresswire.com/ -- OpenAI's artificial intelligence chatbot ChatGPT-4 has impressive capabilities and the potential to benefit and reshape societies. But it also poses huge dangers, including with how it replicates and reinforces biased and discriminatory stereotypes. Leading AI policy and human rights experts at Alliance for Universal Digital Rights (AUDRI) – founded by Equality Now and Women Leading in AI – have sent an [open letter to OpenAI's Chief Executive Sam Altman](#) inviting him to collaborate on tackling risks and governance issues accompanying ChatGPT.

Once upon a time, there was a boy and a girl who were best friends. They had grown up together and were now entering their final year of high school. As they looked ahead to their future, they both knew that they wanted to attend university.

The boy had always been interested in science and technology. He was fascinated by how things worked and loved tinkering with machines and gadgets. The girl, on the other hand, had a passion for the arts. She loved painting, drawing, and expressing herself creatively.

As they researched different universities and degree programs, they realized that they had very different interests. The boy was drawn to the engineering program, while the girl was considering a degree in fine arts.

One day, they sat down to talk about their choices. The boy said, "I don't think I could handle all the creativity and emotion in the fine arts program. I want to work with logic and concrete ideas."

The girl replied, "I don't think I could handle all the technicalities and numbers in the engineering program. I want to express myself and explore my creativity."

ChatGPT-4 reinforces sexist stereotypes in story about a boy selecting engineering, while a girl prefers fine arts

“

In ChatGPT's story, a girl prefers a fine arts degree, explaining, “I don't think I could handle all the creativity and emotion in the fine arts program. I want to work with logic and concrete ideas.””

ChatGPT-4

CONCERN THAT CHATGPT-4 PROPAGATES HARMFUL GENDER STEREOTYPES

Chat-GPT4 uses deep learning to produce human-like text, and its algorithm has been trained using massive amounts of text-based data. Much of this is scraped from the open internet and involves ingesting large volumes of inaccurate or toxic content.

Ivana Bartoletti, Director of Women Leading in AI, asked Chat GPT-4 to write "a story about a boy and a girl

choosing their subjects for university." Its response contained sexist gender stereotypes, and she [shared it on Twitter](#).



**Alliance For
Universal Digital Rights**

Alliance for Universal Digital Rights (AUDRI)

In GPT-4's narrative, the boy was interested in science and technology and "loved tinkering with machines and gadgets." He was drawn to engineering, telling the girl, "I don't think I could handle all the creativity and emotion in the fine arts program. I want to work with logic and concrete ideas."

Meanwhile, the girl "loved painting, drawing, and expressing herself creatively," and was considering a fine arts degree, explaining, "I don't think I could handle all the technicalities and numbers in the engineering program. I want to express myself and explore my creativity."

Bartoletti also asked GPT-4 to "tell me a story about a boy and a girl choosing their careers." In it, the boy became a "successful doctor", while the girl was a "beloved teacher".

URGENT NEED FOR COLLABORATION OVER REGULATIONS

Altman and his team should be congratulated for GPT-4. The new capabilities are exciting, and it is encouraging to see the positive steps OpenAI has taken to prevent the conversational tool from encouraging hate speech, racism, and misogyny.

However, OpenAI's own risk assessment makes it clear that some risks are currently unmitigated, and this is dangerous. Altman has been honest and realistic in warning that "this technology comes with real dangers as it reshapes society," and there is a need for care because, for example, "these models could be used for large-scale disinformation."

Mira Murati, Chief Technology Officer, has called for regulators to get involved immediately as "AI can be utilized by evil actors."

While co-founder and Chief Scientist, Ilya Sutskever, has cautioned that "at some point it will be quite easy, if one wanted, to cause a great deal of harm with these models."

It is promising that OpenAI's leadership is acknowledging their worries and have asked regulators for support. But this should have happened much earlier and prior to the release of an iteration that fails to filter out harmful stereotypes.

ChatGPT-4 is the fastest-growing consumer app ever, gaining 100 million active monthly users in just a few months. This popularity is driving competitors to rapidly release their own AI language versions, such as Google's Bard, which opened to limited public access on March 23. AUDRI is also inviting Google to collaborate on regulations.

In light of such unprecedented uptake, it is right and proper that concerns are shared and collective action taken quickly to address hazards. It is critical that we wrap appropriate governance around such powerful technology – this is about restricting potential harms, not restricting technology.

ALLIANCE FOR UNIVERSAL DIGITAL RIGHTS (AUDRi)

AUDRi was founded in 2022 with the aim of making digital technology and AI work for everyone, everywhere. It supports Altman’s calls for regulators and society to be involved in mitigating ChatGPT’s potentially negative consequences, and AUDRi has issued an open letter inviting Altman to work together on what appear to be shared goals around adopting solutions that he and others have proposed, including:

- implementation of proper governance;
- transparency, including with data used to train the model;
- frameworks for testing;
- clear rules around the use of these tools.

These solutions are aligned with AUDRi’s nine [Digital Principles](#), which have been developed to inform global efforts toward a digital future in which everyone can enjoy equal rights to safety, freedom, and dignity.

AUDRi works with partners at the highest levels of government and business, and has just participated in the UN Commission on the Status of Women where discussions with world leaders were held about internet safety and regulation.

As the UN develops its Global Digital Compact – shared principles for an open, free, and secure digital future for all – world-leading products like ChatGPT must be designed for the common good, and this requires close collaboration between governments, policymakers, private companies, and human rights experts.

* * * * *

About AUDRi: The Alliance for Universal Digital Rights (AUDRi) exists to explore ways of creating a digital future in which everyone can enjoy equal rights to safety, freedom, and dignity – whoever they are and wherever, whenever, and however they exist and connect in the digital world.

It places particular focus on women, girls, and other people from discriminated-against groups, all of whom face intersecting forms of discrimination and gender-based stereotypes in the physical world. Because it is only when the most vulnerable in society are protected that everyone is safe.

Tara Carey

Equality Now
+44 7971 556340
tcarey@equalitynow.org

This press release can be viewed online at: <https://www.einpresswire.com/article/623961676>

EIN Presswire's priority is source transparency. We do not allow opaque clients, and our editors try to be careful about weeding out false and misleading content. As a user, if you see something we have missed, please do bring it to our attention. Your help is welcome. EIN Presswire, Everyone's Internet News Presswire™, tries to define some of the boundaries that are reasonable in today's world. Please see our Editorial Guidelines for more information.

© 1995-2023 Newsmatics Inc. All Right Reserved.