AI platform allegedly bans journalist over fake Trump arrest images

Enlarge / AI-generated photo faking Donald Trump’s possible arrest, created by Eliot Higgins using Midjourney v5.

reader comments

94 with

Share this story

Yesterday, about 50 AI-generated images imagining what former President Donald Trump’s arrest would look like spread across Twitter. The images caused confusion for some users because the Midjourney v5 engine used to create them generates such realistic-looking content.

Today, the Midjourney user behind the images—Bellingcat founder Eliot Higgins—alleged to Buzzfeed News that he was banned from Midjourney over the images. Buzzfeed also reported that Midjourney has banned the word “arrest” from being used in text prompts to generate images.

Midjourney has not yet confirmed the ban. An Ars test indicates that prompts using Donald Trump’s name also seem to be blocked on the engine, generating instead an error message saying that the prompt was blocked because “it may conflict with our content policy.” We got the same error message when we used prompts including other Trump family names and all the presidents’ names. However, other celebrity names, like former NFL player Tom Brady, are still seemingly acceptable in

prompts.

Higgins told Buzzfeed that he understands why Midjourney may have banned him, saying that “it was pushing my luck” creating the Twitter thread that went viral. Higgins told Ars that responses to the images have been generally positive but that he also is “getting quite a few angry” Trump fans responding.

Advertisement

“I didn’t intend to do any clever criticism or anything like that,” Higgins told Buzzfeed, saying that he was just having fun imagining Trump’s arrest with the images. “But then it kind of took on a life of its own.”

Higgins told Ars that “what really stood out” to him was “the number of people who don’t seem to have been aware of Midjourney’s existence or what it’s capable of.” He thinks creating the images “had a positive impact in raising awareness of what kind of images can be faked now” and said that AI platforms like Midjourney “probably shouldn’t train their datasets on people they don’t want re-created.”

Yesterday, Higgins told Wired how Twitter users could avoid being fooled by convincing AI-generated images like the fake Trump arrest images. Higgins noted that celebrities are easier targets for deepfakes, though, because “it’s clear that the

more famous a person is, the more images the AI has had to learn from," Higgins said.

"So very famous people are rendered extremely well, while less famous people are usually a bit wonky," Higgins told Wired.

On Twitter, Higgins' images are now labeled with a community note flagging that they're fake. Twitter's policy states that users "may not deceptively share synthetic or manipulated media that are likely to cause harm" and that these labels "help people understand their authenticity" and "provide additional context."

Higgins' entire thread is still available on Twitter, where some images have been viewed 4.8 million times as of this writing.

While Trump previously indicated that he expected his indictment to occur yesterday, that didn't happen. Outlets like MSNBC and The New York Times are now reporting that his potential arrest over hush-money payments to porn star Stormy Daniels could happen in the coming days.

While Ars has not found any comments from Trump on Higgins' AI-generated images, Forbes reported that Trump has told advisors that he wants his arrest—if it happens—to be a "spectacle" that includes a "perp walk" in handcuffs so that he doesn't appear weak to his supporters.

https://arstechnica.com/?p=1925983