Fake Pentagon “explosion” photo sows confusion on Twitter



Enlarge / A fake AI-generated image of an “explosion” near the Pentagon that went viral on Twitter.

reader comments

42 with

On Monday, a tweeted AI-generated image suggesting a large explosion at the Pentagon led to brief confusion, which included a reported small drop in the stock market. It originated from a verified Twitter account named “Bloomberg Feed,” unaffiliated with the well-known Bloomberg media company, and was quickly exposed as a hoax. However, before it was debunked, large accounts such as Russia Today had already spread the misinformation, The Washington Post reported.

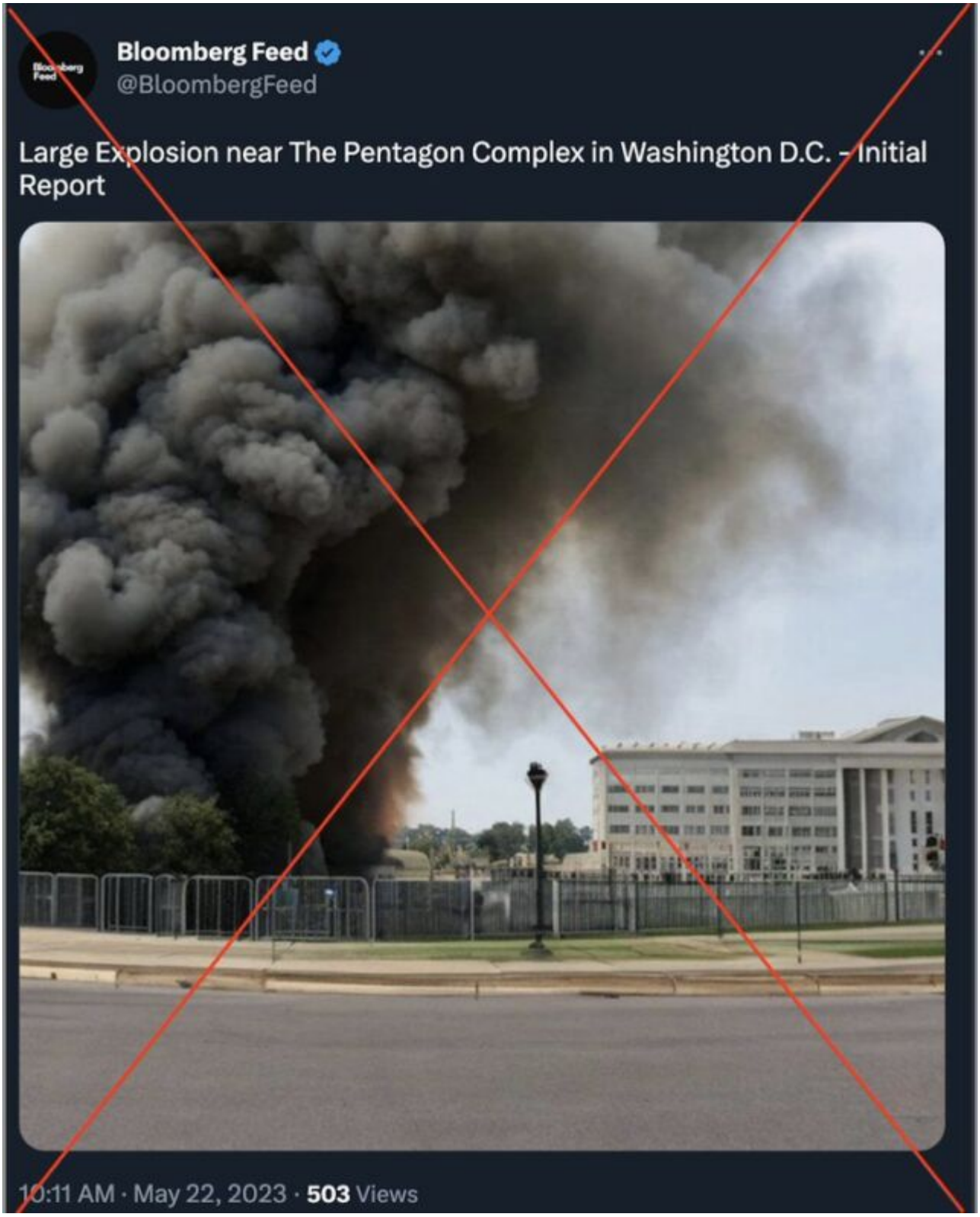
The fake image depicted a large plume of black smoke alongside

a building vaguely reminiscent of the Pentagon with the tweet “Large Explosion near The Pentagon Complex in Washington D.C. – Inital Report.” Upon closer inspection, local authorities confirmed that the image was not an accurate representation of the Pentagon. Also, with blurry fence bars and building columns, it looks like a fairly sloppy AI-generated image created by a model like Stable Diffusion.

Before Twitter suspended the false Bloomberg account, it had tweeted 224,000 times and reached fewer than 1,000 followers, according to the Post, but it’s unclear who ran it or the motives behind sharing the false image. In addition to Bloomberg Feed, other accounts that shared the false report include “Walter Bloomberg” and “Breaking Market News,” both unaffiliated with the real Bloomberg organization.

This incident underlines the potential threats AI-generated images may present in the realm of hastily shared social media—and a paid verification system on Twitter. In March, fake images of Donald Trump’s arrest created with Midjourney reached a wide audience. While clearly marked as fake, they sparked fears of mistaking them for real photos due to their realism. That same month, AI-generated images of Pope Francis in a white coat fooled many who saw them on social media.

Advertisement



Enlarge / A screenshot of the “Bloomberg Feed” tweet about the reported explosion near the Pentagon that was later confirmed to be fake.

The pope in puffy coats is one thing, but when someone features a government subject like the headquarters of the United States Department of Defense in a fake tweet, the consequences could potentially be more severe. Aside from

general confusion on Twitter, the deceptive tweet may have affected the stock market. The Washington Post says that the Dow Jones Industrial Index dropped 85 points in four minutes after the tweet spread but rebounded quickly.

Much of the confusion over the false tweet may have been made possible by changes at Twitter under its new owner, Elon Musk. Musk fired content moderation teams shortly after his takeover and largely automated the account verification process, transitioning it to a system where anyone can pay to have a blue check mark. Critics argue that practice makes the platform more susceptible to misinformation.

While authorities easily picked out the explosion photo as a fake due to inaccuracies, the presence of image synthesis models like Midjourney and Stable Diffusion means it no longer takes artistic skill to create convincing fakes, lowering the barriers to entry and opening the door to potentially automated misinformation machines. The ease of creating fakes, coupled with the viral nature of a platform like Twitter, means that false information can spread faster than it can be fact-checked.

But in this case, the image did not need to be high quality to make an impact. Sam Gregory, the executive director of the human rights organization Witness, pointed out to The Washington Post that when people want to believe, they let down their guard and fail to look into the veracity of the information before sharing it. He described the false Pentagon image as a “shallow fake” (as opposed to a more convincing “deepfake“).

“The way people are exposed to these shallow fakes, it doesn’t require something to look exactly like something else for it to get attention,” he said. “People will readily take and share things that don’t look exactly right but feel right.”

https://arstechnica.com/?p=1941475