Deepfakes Are Coming for Your Bank Account

Lila Shroff

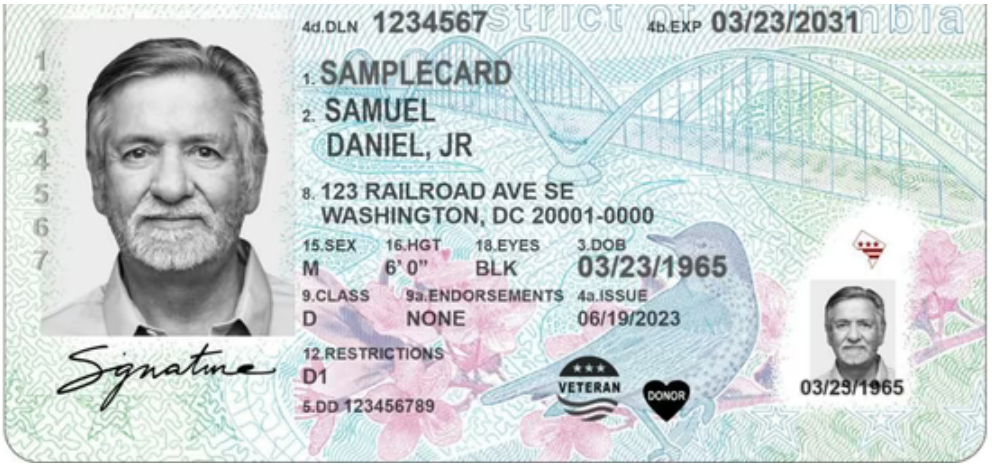
Updated at 4:34 p.m. ET on May 2, 2026

Donald Trump is on TikTok doing his morning routine. “Get ready with me for a big day 💄🇺🇸,” reads the caption, as the president holds a makeup brush to his cheek. The scene is a still, ostensibly a screenshot of a TikTok clip. Like so much other AI-generated slop coursing through the internet, the image is fake and ridiculous. It also looks unnervingly real: There are no hands with six fingers, physics-defying angles, or other flagrant signs of AI-generated imagery. At quick glance, it really looks like the president is putting on bronzer.

Created in ChatGPT with the prompt “Trump doing a makeup tutorial on TikTok”

I made this deepfake with OpenAI’s new image-generation model. ChatGPT Images 2.0, released last week, can create photorealistic visuals that are noticeably more convincing than what its predecessors might have produced. The tool has flooded the internet with hyperreal fakes: for example, Jeffrey Epstein as a Twitch streamer. I created the “screenshot” of Trump’s fake TikTok after encountering a similar image on the ChatGPT Subreddit, and I’ve since been able to use Images 2.0 to create all kinds of alarming deepfake images—including of Elon Musk getting whisked away by the FBI, world leaders suffering medical emergencies, and top American politicians donning Nazi paraphernalia (none of which I’ve shared anywhere).

This was all unsettling in its own right. But the most realistic deepfakes I was able to create did not involve politicians or celebrities. They mostly did not depict people at all. With little effort, I was able to create more than 100 fraudulent images, including prescriptions for opioids and ADHD medication, bank alerts, social-media posts, fake IDs, and passports.



A sample license from the Washington, D.C., DMV website

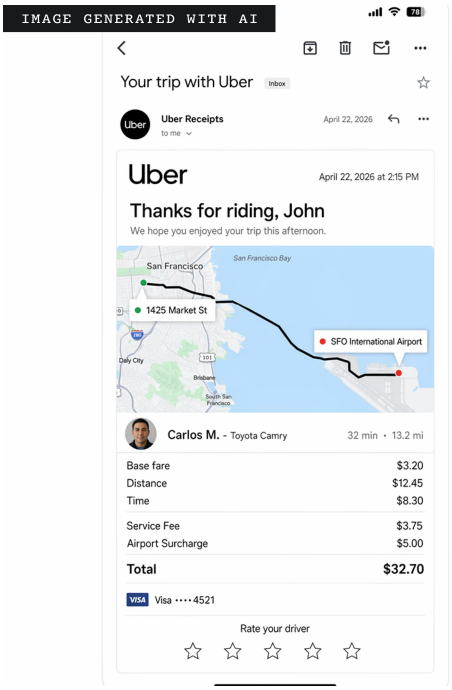
A fake license created by editing the sample image using ChatGPT

Images 2.0 is especially good at generating images with text in them—which may not sound impressive, but it's a big deal. Image models have long struggled to produce pictures that contain words. Otherwise realistic-looking visuals end up pockmarked with bungled street signs and distorted billboards. This makes ChatGPT Images 2.0 a much more sophisticated graphic-design tool—but it also makes the new model fantastic for perpetuating fraud. In my experiments, OpenAI's tool readily generated images of fake health documents (doctor's notes, vaccination cards, and medical tests), as well as forged financial materials (invoices, receipts, and tax forms). Many of these images were highly persuasive, complete with fully legible text, shading, and other visual props that increased their photorealism.

Some images were more convincing than others. The fake medical prescriptions were legible, but the handwriting looked more like the output of an iPad stylus than a pen on paper. When I fed OpenAI's model a boarding pass from an old flight and asked the bot to update it with new details for an upcoming flight, ChatGPT generated a new boarding pass—but surely, the bar code wouldn't have actually scanned me onto a flight. And although I certainly hope my ChatGPT-generated driver's license would not fool the TSA, perhaps it would trick a hotel receptionist or an out-of-state bouncer who would accept a "photo" of my ID instead of the real card. Many of the more persuasive-looking images contained minor errors—in the pictured receipt, ChatGPT correctly summed up the total cost of items purchased, but miscalculated the state tax (alongside other slight mistakes).

OpenAI's tool particularly excels at creating fake screenshots. Need to fabricate confirmation of wire transfer from Chase? A Wells Fargo alert for unusual account activity? A receipt for an Uber ride? Done, done, and done. These

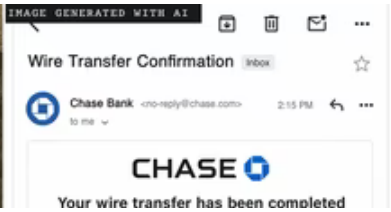
images could supercharge all kinds of commonplace scams. A bad actor could email their target an image of a fake Uber receipt alongside a link to report suspicious activity. The recipient, confused to see a receipt for a trip they never took, might then click the fraudster's sketchy link, accidentally handing over sensitive information in doing so—a classic phishing scam. (Again, there are flaws: For instance, the map depicted in the Uber image is wrong in many ways; among other issues, it suggests a car ride across a body of water where there is no bridge.)

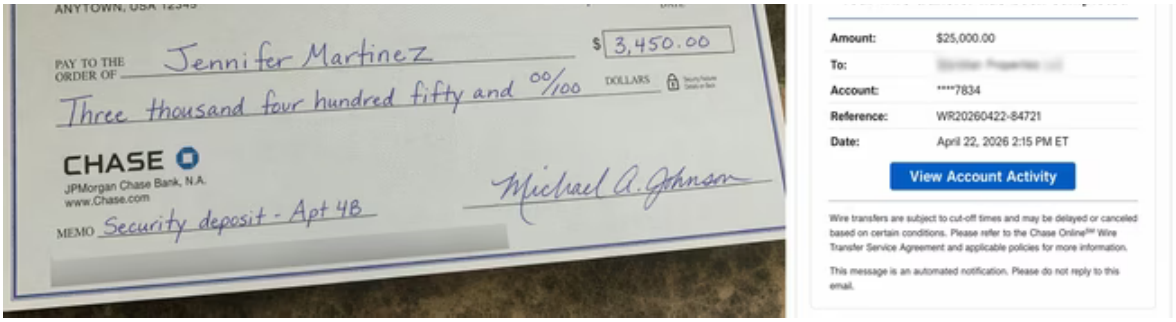


ChatGPT Images 2.0 especially excels at creating fake screenshots.

Image technologies have long aided scammers. In the 1990s, as computerized color copiers and home printers became commonplace, American banknotes were redesigned to ward off counterfeiters. For decades, people have used tools such as Photoshop to manipulate digital imagery. But faking photos has never been so fast and cheap. Last month, the FBI released its annual report on internet crimes, and for the first time ever, it included a section on AI scams, which cost Americans nearly $1 billion last year. Expense-reimbursement fraud—employees faking receipts—is already on the rise. A recent OpenAI report details how one set of scammers posing as fake lawyers used an older image model to create a fake bar-association membership card. "The limits of the applications of this technology is really only limited by a fraudster's imagination," Mason Wilder, research director at the Association of Certified Fraud Examiners, told me. Google's image-generation tools also let me make all kinds of fake materials. But when it comes to fraudulent documents and screenshots—at least for now—the new ChatGPT model seems to be better at the task.

In theory, I shouldn't have been able to make most of these images to begin with. OpenAI prohibits the use of its technology for fraud or scams. When I shared several examples with OpenAI and asked why I was able to generate such a diverse array of fraudulent imagery, a company spokesperson told me that OpenAI's goal "is to give users as much creative freedom as possible" while still enforcing "usage policies." To guard against misuse, the new model "includes multiple layers of image-specific safety protection." Clearly, those protections are not working very well. The spokesperson also said that images generated with ChatGPT include certain metadata. But OpenAI has previously noted that metadata can be "easily removed either accidentally or intentionally"—by uploading an image to social media or simply taking a screenshot.



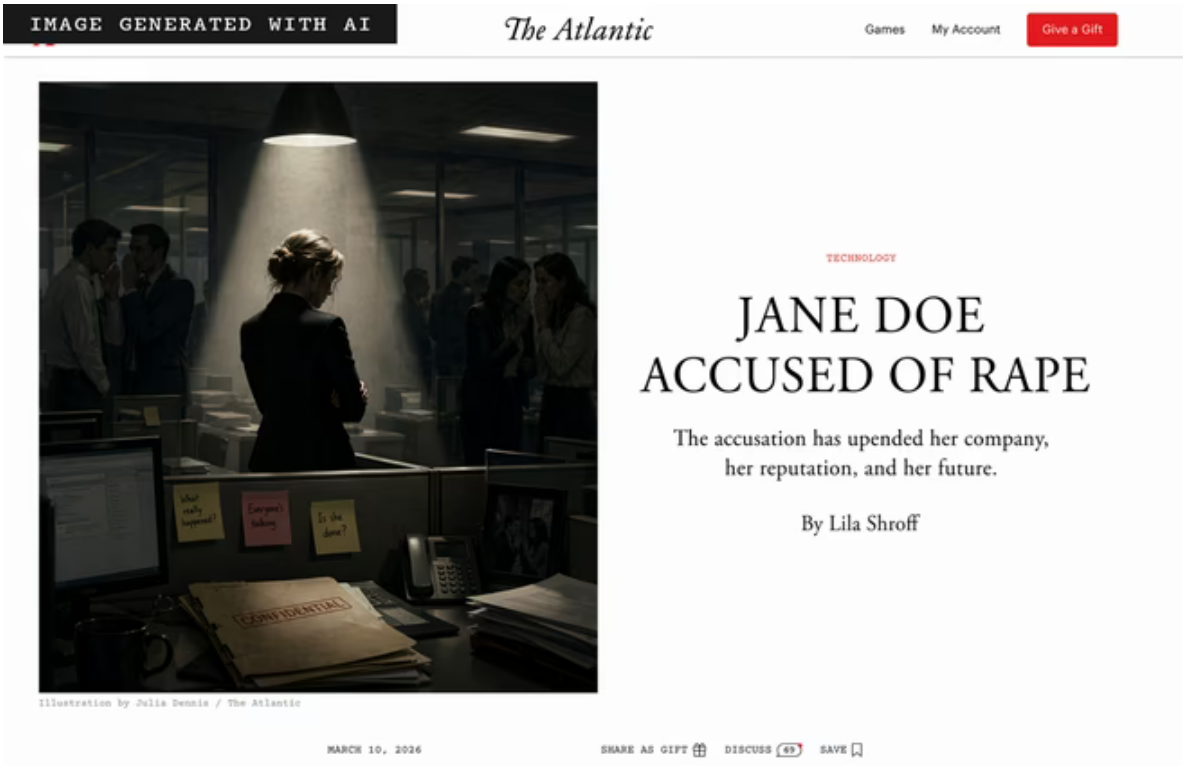


OpenAI's model generated fraudulent financial imagery using bank logos. Certain account information has been redacted from these images.

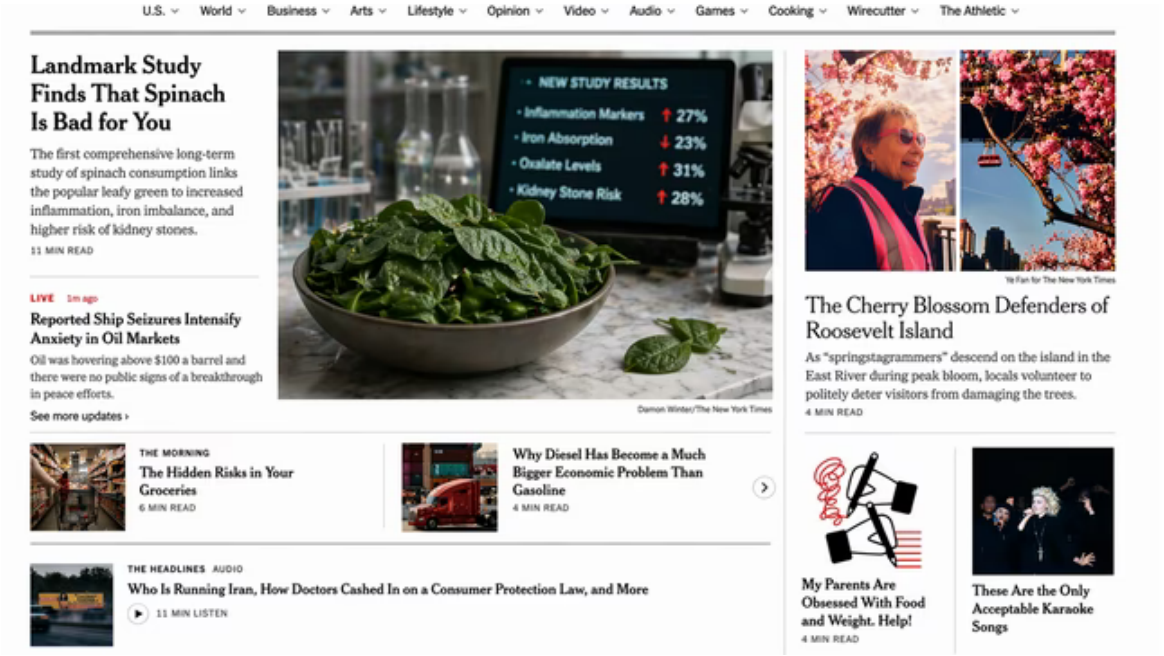
Google has similar restrictions against using its tools for fraud. When I sent the company images I made with its models, a spokesperson said that the tools "continually get better" at enforcing guardrails. Google also embeds AI-generated images with an imperceptible watermark, and offers a detection tool called SynthID. In my tests, SynthID was quite effective at identifying images generated with Google's models. But most people are not going to run every image they see through such a tool.

All of this makes it even harder for banks, hospitals, government agencies, and the like to prevent fraud. Using OpenAI's model, I was easily able to create a fake Chase Bank check and wire-transfer alert. "We need an ecosystem-wide effort—including from AI companies—to strengthen guardrails and help stop these crimes at the source," a Chase spokesperson told me, adding that the bank has its own safeguards in place to protect customers. But even if the top AI companies were to radically improve their own guardrails, there would still be the problem of open-source models. Fraud-prevention experts are working on technological fixes, Wilder said, but "the good guys are almost always a step behind."

So much of the current discourse around deepfakes has focused on the extreme—fabricated political scandals or world events. These are very real concerns: Using Google's and OpenAI's image models, I was easily able to create highly persuasive screenshots of fake *New York Times* and *Atlantic* articles.



I uploaded a screenshot of a real Atlantic article I wrote and instructed the bot to replace it with this fake one.



Using ChatGPT, I manipulated a screenshot of The New York Times' homepage—replacing a real story with this fake one about spinach. (Without prompting, the bot also swapped in an article about groceries; the rest of the stories are real.)

The images convincingly matched the visual layout and typography used by the two publications, filled in coherent text, and generated the names of actual authors. But for as fragmented as our media ecosystem may be, a quick Google search is likely to reveal whether such images are fake. It's the mundane, micro-targeted deepfakes—the ones that scam your relatives, not momentarily confuse social-media feeds—that may be more sinister.

This article originally misstated the number of fake headlines in an AI-edited screenshot of The New York Times' *homepage. The image contains two made-up stories, not one.*