



How to Tell If a Photo Is an AI-Generated Fake

怎样识别 AI 生成的假图片

Date: Mar. 31, 2023

You may have seen photographs that suggest **otherwise**, but the pope didn't wear a stylish, brilliant white **puffer coat**. These recent viral hits were the fruits of artificial intelligence systems that process a user's textual **prompt** to create images. They demonstrate how these programs have become very good very quickly—and are now **convincing** enough to fool an **unwitting** observer.

罗马教皇方济各没有穿过一件时尚亮眼的白色羽绒服，而你可能看到过与之相悖的照片。最近这些爆火的图片是人工智能系统的成果，它们处理用户的文字提示以创建图像。这些图片证明了这些程序是如何在极短时间内变得非常强大，并且它们现在已经足够令人信服，可以欺骗不知情的观众。

So how can skeptical viewers spot images that may have been generated by an artificial intelligence system such as DALL-E, Midjourney or Stable Diffusion? Each AI image generator—and each image from any given generator—varies in how convincing it may be and in what **telltale** signs might **give** its **algorithm** away. For instance, AI systems have historically struggled to **mimic** human hands and have produced **mangled appendages** with too many **digits**. As the technology improves, however, systems such as Midjourney V5 seem to have **cracked** the problem—at least in some examples. **Across the board**, experts say that the best images from the best generators are difficult, if not impossible, to **distinguish** from real images.

那么，持怀疑态度的观众如何才能看出可能是由 DALL-E、Midjourney 或 Stable Diffusion 等人工智能系统生成的图像呢？每个人工智能图像生成

617 Words

otherwise ['ʌðərwaɪz] *adv.* 以其他方式；除此以外；否则，不然 | You know what this is about. Why pretend otherwise? 你明明知道这是怎么回事，为什么装作不知道？

puffer coat 羽绒服

prompt [prɒːmpt] *n.* 提示词，提示 *v.* 促使，导致；提示，提醒 *adj.* 立即的，迅速的 | A message will appear which will prompt you for certain information. 会有信息显示，提醒你输入某些资料。

convincing [kən'vɪnsɪŋ] *adj.* 令人信服的，有说服力的 | She sounded very convincing to me. 我觉得她的话很有说服力。

unwitting [ʌn'wɪtɪŋ] *adj.* 不知情的，无意的 | She was the unwitting cause of the argument. 她无意中引起了这场争执。

telltale ['telteɪl] *adj.* 暴露实情的，能说明问题的 | They examined the child carefully, looking for telltale signs of abuse. 他们仔细检查了那个孩子，寻找被虐待的痕迹。

give ... away 泄露，暴露 | The examiners have given away the answer. 考官们泄露了答案。

algorithm ['ælgərɪðəm] *n.* 算法

器——以及来自任何特定生成器的每张图像——在令人信服的程度以及可能泄露其算法的破绽方面各不相同。例如，人工智能系统一直在努力模仿人类的手，生成的图片中手部畸形，手指过多。然而，随着技术的进步，像 Midjourney V5 这样的系统似乎已经解决了这个问题——至少在某些例子中是这样。总的来说，专家们表示，要想将最好的生成器生成的最佳图像与真实图像区分开来，即使不是不可能，也是很难做到的。

One recent experiment highlighted how well AI is able to deceive. Sophie Nightingale, a psychologist at Lancaster University in England who focuses on digital technology, co-authored a study that tested whether online volunteers could distinguish between passportlike **headshots** created by an AI system called StyleGAN2 and real images. The results were **disheartening**, even back in late 2021, when the researchers ran the experiment.

最近的一项实验凸显了人工智能的欺骗能力有多强。英国兰卡斯特大学专攻数字技术的心理学家索菲·南丁格尔参与了一项研究，该研究测试了网络志愿者能否区分由名为 StyleGAN2 的人工智能系统创建的类似护照照片的头像和真实图像。即使是在研究人员开展实验的 2021 年底，结果也令人十分沮丧。

The researchers tried to help the test subjects improve their AI-detecting abilities. They prepared participants in advance by having them read through advice for detecting artificially generated images. That advice highlighted areas where AI algorithms often **stumble** and create mismatched earrings, for example, or **blur** a person's teeth together. Nightingale also notes that algorithms often struggle to create anything more **sophisticated** than a plain background. But even with these additions, participants' accuracy only increased by about 10 percent, she says—and the AI system that generated the images used in the **trial** has since been upgraded to a new and improved version.

研究人员试图帮助测试对象提高他们察觉人工智能的能力。他们提前
关注公众号：拾点外刊 免费下载电子版 附音频

mimic ['mɪmɪk] v. 模仿 | He mimicked her upper-class accent. 他模仿她那上流社会的腔调。

mangled ['mæŋɡld] adj. 面目全非的，严重损毁的

appendage [ə'pendɪdʒ] n. 附加物，附属物

digit ['dɪdʒɪt] n. 手指，拇指，脚趾；数字，数位

crack [kræk] v. 找到解决（难题等的）方法；破裂，裂开 | It took them nearly two months to crack the code. 他们花了将近两个月的时间才破译这个密码。

across the board 全体，整体 | The industry needs more investment across the board. 这一行业需要全面增加投资。

distinguish [dɪ'stɪŋɡwɪʃ] v. 区分，辨别 | Could he distinguish right from wrong? 他能分辨是非吗？

headshot [hed'ʃɒt] n. 大头照，头像

disheartening [dɪs'hɑ:tnɪŋ] adj. 令人沮丧的，让人泄气的 | It's disheartening to see what little progress has been made. 看到没取得什么进展令人感到灰心丧气。

stumble ['stʌmbl] v. 出岔子，出差错；跌跌撞撞地走，踉跄；结巴 | There are plenty of young rivals to take his place if he stumbles. 如果他出了差错，有许多年轻的竞争对手会取而代之。

blur [blɜ:r] v. （使）模糊，（使）看

让参与者阅读有关检测人工生成图像的建议。该建议强调了人工智能算法经常出错的地方，例如耳环不配对，牙齿模糊成一片。南丁格尔还指出，算法往往很难创造出较为复杂的背景。但她说，即使补充了这些技巧，参与者的准确率也只提高了 10% 左右，而且试验中使用的生成图像的人工智能系统此后已经升级到了一个改良的新版本。

Ironically, as image-generating technology continues to improve, humans' best defense from being fooled by an AI system may be yet another AI system: one trained to detect artificial images. Experts say that as AI image generation progresses, algorithms are better equipped than humans to detect some of the tiny, **pixel-scale fingerprints** of robotic creation.

具有讽刺意味的是，随着图像生成技术的不断改进，人类避免被人工智能系统愚弄的最佳防御手段可能是另一个人工智能系统：一个经过训练可以检测人工图像的系统。专家表示，随着人工智能图像生成技术的发展，算法比人类更有能力检测到机器创造的一些微小的、像素级的破绽。

Still, these systems have significant **shortcomings**, experts say. Most such algorithms are trained on images from a specific AI generator and are unable to identify fakes produced by different algorithms. Most detectors also lack the user-friendly **interfaces** that have **tempted** so many people to try the generative AI systems.

尽管如此，专家表示，这些检测系统存在重大缺陷。大多数这样的算法都是根据来自特定人工智能生成器的图像进行训练的，它们无法识别由不同算法产生的假图像。生成人工智能系统的用户界面更友好，吸引了许多人尝试，而大多数图像检测系统则缺乏这样用户友好的界面。

Moreover AI detectors will always be **scrambling to keep up with** AI image generators, some of which incorporate similar detection algorithms but use them as a way to learn how to make their fake output less detectable. "The battle between AI systems that generate images and AI systems that detect the AI-generated images is going to be an **arms race**," says Wael

不清；（使）难以区分 *n.* 模糊不清的事物 | Tears blurred her eyes. 泪水模糊了她的视线。

sophisticated [sə'fɪstɪkeɪtɪd] *adj.* 复杂的，精密的；富有经验的，老练的 | Medical techniques are becoming more sophisticated all the time. 医疗技术日益复杂精妙。

trial ['traɪəl] *n.* 试验，试用；审理，审判 | The new drug is undergoing clinical trials. 这种新药正在进行临床试验。

ironically [aɪ'rɒnɪkli] *adv.* 具有讽刺意味的是，令人哭笑不得的是 | Ironically, his cold got better on the last day of his holiday. 具有讽刺意味的是，他的感冒竟在假期的最后一天有所好转。

pixel ['pɪksl] *n.* 像素（计算机屏幕上图像最小的单元）

fingerprint ['fɪŋɡəprɪnt] *n.* 指纹；与众不同的特征，（鉴别某种特征的）依据 | The author's fingerprint is quite obvious in all of his works. 作者的特色在其全部作品中显得十分突出。

shortcoming ['ʃɔ:rtkʌmɪŋ] *n.* 缺点，短处

interface ['ɪntərfeɪs] *n.* （软件的用户）界面；（计算机设备之间的）连接，接口 | the user interface 用户界面

tempt [tempt] *v.* 诱惑，吸引 | Don't be tempted to spend too much. 别抵制不了诱惑花太多的钱。

AbdAlmageed, a research associate professor of computer science at the University of Southern California. “I don’t see any side winning anytime soon.”

此外，人工智能检测器将不停追赶人工智能图像生成器的步伐，一些图像生成器包含类似的检测算法，但将它们用作学习如何使产出的假图片更不容易被检测到。“人工智能图像生成系统和检测系统之间的战斗将是一场军备竞赛，”南加州大学计算机科学研究副教授韦尔·阿卜杜勒·马吉德说。“我认为二者很难在短时间内分出胜负。”

AbdAlmageed says no approach will ever be able to catch every single artificially produced image—but that doesn’t mean we should give up. He suggests that social media platforms need to begin **confronting** AI-generated content on their sites because these companies are better posed to **implement** detection algorithms than individual users are.

阿卜杜勒·马吉德说，没有一种方法能够捕捉到每一张人工生成的图像，但这并不意味着我们应该放弃。他建议社交媒体平台需要开始在其网站上处理人工智能生成的内容，因为这些公司比个人用户更有能力运用检测算法。

And users need to more skeptically **evaluate** visual information by asking whether it’s false, AI-generated or harmful before sharing. “We as human species sort of grow up thinking that **seeing is believing**,” AbdAlmageed says. “That’s not true anymore. Seeing is not believing anymore.”

用户需要抱着更大的怀疑态度去评估视觉信息，在分享之前问一问它是否是虚假的、人工智能生成的或是有害的。阿卜杜勒·马吉德说：“作为人类，我们从小就认为眼见为实。但现在不是这样了，眼见并不一定为实了。”

scramble [ˈskræmbəl] v. 仓促完成；争抢；攀登 | They were scrambling to give the impression that the situation was under control. 他们急于给人一种形势已得到控制的印象。

keep up with 齐步前进，跟上；熟悉，了解（消息、形势等） | Wages are not keeping up with inflation. 工资赶不上通货膨胀。

arms race 军备竞赛

confront [kənˈfrʌnt] v. 处理，解决（问题或困境）；面对，面临 | She knew that she had to confront her fears. 她心里明白自己必须克服恐惧心理。

implement [ˈɪmplɪmənt] v. 贯彻，执行，实施 n. 用具，器具 | We have decided to implement the committee’s recommendations in full. 我们已决定全面实施委员会提出的建议。

evaluate [ɪˈvæljuːeɪt] v. 评价，评估 | The performance of each employee is evaluated once a year. 每年要对每位雇员的表现进行一次评估。

seeing is believing 眼见为实

➤ 脉络梳理

- Para. 1 指出如今人工智能生成的图片足以以假乱真，在网络上广泛传播
- Para. 2-4 举例说明人类很难区别真实的图片和人工智能生成的图片
- Para. 5-7 指出识别人工智能生成的图片的最佳方式是人工智能图片检测系统，但其存在缺陷
- Para. 8-9 提出建议：社交媒体平台应主动处理人工智能生成的内容，个人在分享之前应确认图片是否真实

➤ 长难句拆解

