

RUBIN

SPECIAL ISSUE

SCIENCE MAGAZINE

IT SECURITY

Three tough nuts for quantum computers to crack

This is how artificially generated images reveal their true colours

Start-up: Ready for the new generation of mobile communications

```
def generate(prompt, num_images=4):
    prompt_list = [prompt] * num_images

    with autocast("cuda"):
        images = pipe(prompt_list).i

    for i, image in enumerate(images):
        image.save(f"images/{prompt}_{i}.png")

for _ in range(25):
    generate("hyper realistic and
```

In the background information of images clues can be found that indicate that the image was artificially created. (photo: ms)

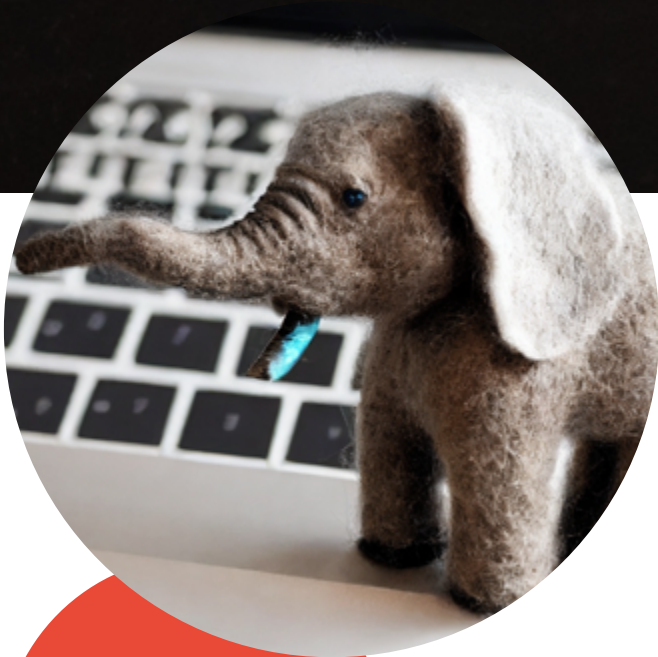
Humans often have no chance whatsoever of distinguishing artificially created images, audio or videos from the real deal.

This is why researchers of the Horst Görtz Institute for IT Security are working on automated recognition

Vladimir Putin stands behind a lectern and addresses the USA: he has very much the means to undermine democracy in the USA – but he claims that he doesn't have to. The US would take care of that themselves. Society is divided already. The video looks real – but it is not. Youtube is flooded with such clips, some of which are quite well done, some of which are not. "It's still a lot of work, but if you want to, you can, for example, superimpose the face of a famous person on the body of another person so skilfully that viewers won't notice it at first glance," says Jonas Ricker.

For his doctoral thesis, which he is writing at the Faculty of Computer Science, he has specialised in fake images. The focus of his work, however, is not videos but photos. He can whip up several links at the drop of a hat that will show you pictures of people who don't exist or where you can try

THIS IS HOW ARTIFICIALLY GENERATED IMAGES REVEAL THEIR TRUE COLOURS



Looks like the real thing: this wool elephant was created by text-to-image generation. (photo: Hugging Face)

to guess whether the picture of the depicted person is real or not. The fake images are generated using deep learning, a machine learning method – hence the name “deep fake”. “When older methods are used, you can sometimes spot anomalies in the symmetry,” he points out. “For example, different earrings will be a telltale sign, as will asymmetrical glasses. But the methods are getting better and better, and studies have proven that people tend to be rather bad at distinguishing real images from fake ones.”

One process for generating such images is called GAN, short for generative adversarial networks. “Basically, such networks are always divided into two parts: one part generates the image, another, the so-called discriminator, decides whether the generated image looks real or not,” Jonas Ricker illustrates. “Picture it like this: one part is a counterfeiter, the oth- ▶



er part is the police who have to tell fake banknotes from real ones.” The artificial intelligence makes this decision on the basis of many real images that are fed in as a learning dataset. At first, the generator merely generates any random pixels. As it progresses, it learns more and more through feedback from the discriminator. The discriminator also gets better and better at distinguishing the generator’s images from real ones. The generator and discriminator train each other, so to speak, which ultimately results in images that look deceptively real.

In an article published in 2020, Jonas Ricker’s former colleague Joel Frank describes a way of detecting fake images. The key lies in the so-called frequencies. “It’s difficult to explain what frequencies are in images,” says the researcher. The best way is to think of frequencies as light-dark differences. Low frequencies are common in people’s faces. High frequencies can be found in hair, for example, and they are perceived at a more subconscious level. Consequently, an image in which high frequencies have been altered will look al-

most exactly the same to us as the original image. However, technology is not so easily fooled: “When it comes to high frequencies, GAN-generated images show characteristic deviations from real photos,” explains Jonas Ricker. In artificially generated images, high frequencies occur in excess. This is traceable, and it allows the images to be distinguished from real photos.

Jonas Ricker is currently working on another class of models for image generation, the so-called diffusion models. While GANs were already introduced in 2014, diffusion models have only been researched for roughly three years, with outstanding results. “The basic principle of diffusion models sounds surprising at first,” says Ricker: “A real image is destroyed step by step by adding Gaussian noise. After a few hundred steps, no image information is left, the image is completely distorted. The goal of the model is now to reverse this process to reconstruct the original image – which is a difficult problem.”



Artificial intelligences are able to create images that humans cannot distinguish from photographs. (photos: ms)

The key is not to predict the image directly, but to proceed step by step, as with noise. With a sufficiently large amount of training data, the model can learn to make a noisy image a little bit less noisy. By repeating the process again and again, completely new images can then be created from random noise. “One weakness of this method is the long processing time due to the several hundred steps involved,” admits Jonas Ricker. “Still, techniques for optimisation have already been introduced and research is constantly making progress.”

Recently, diffusion models have caused quite a stir with so-called text-to-image generation. This allows images to be generated on the basis of text input – with an astonishing level of detail. These models are trained with the aid of countless image-text pairs sourced from the internet. Both this data collection and the actual training require a lot of computing power and are therefore extremely expensive. Until recently, only large companies like Google (Imagen) and OpenAI (DALL-E 2) were able to train these models in high quality – ▶



”
ULTIMATELY,
ANY IMAGE MAY
BE TREATED
WITH SUSPI-
CION AND CAN
BE POTENTIALLY
DISPUTED, EVEN
IMAGES THAT
ARE USED AS
EVIDENCE IN A
COURT OF LAW.

“

Jonas Ricker

and they keep the models largely under wraps. Today, there's also "stable diffusion", a freely accessible model that anyone can use, provided that their computer has enough power. The requirements are moderate, and websites do exist that allow you to create images for your own texts.


The diffusion model is powered by an organisation that has the necessary resources and computing power thanks to a donation. "The model is already very good at generating deceptively real images and will continue to improve in the future," believes Jonas Ricker. This makes it even more difficult to distinguish real images from those generated in this manner. Here, the frequency approach is already less accurate than it is for GAN images. "Another approach is to use the reflections of light in the eyes in order to tell the difference – this, at least, is possible with pictures of humans," says Jonas Ricker. He's currently testing various approaches that make it possible to distinguish images generated by the model from real photos. A universal detector that works for all types of GAN images doesn't actually work that well for this type of image – unless you fine-tune it to make it more accurate. This means that the detector, which is supplied with a lot of real and fake images as learning material along with the relevant information if they are indeed real or fake, is fed additional training data in order to optimise the detection for the new data. This is how it can learn to correctly tell which images have been generated by the diffusion model. How it does this, however, is unclear.

The ability to distinguish between real and fake images is crucial not only in order to expose fake news, including those in video format, but also to detect fake profiles on social media. Such profiles are used on a large scale, for example to influence public opinion in the political arena. "This is exactly what the CASA Cluster of Excellence aims to do: expose large-scale adversaries such as governments or intelligence agencies that have the resources to use deep fakes to spread propaganda," says Jonas Ricker.

The detection of fake photos is also relevant under criminal law, for example when it comes to unintentional pornography in which people's faces are pasted onto the bodies of others. "Generally speaking, the mass of artificially created images leads to a loss of trust, including the trust in reputable media, points out Jonas Ricker. "Ultimately, any image may thus be treated with suspicion and can be potentially disputed, even images that are used as evidence in a court of law."

Even though Ricker aims to ensure that fake pictures can be detected automatically, he reckons that it will ultimately come down to something else entirely: "I think in the end of the day genuine pictures will have to be certified," he speculates. "A feasible approach might be to use cryptographic methods, which would have to be integrated in the photographer's camera, making every genuine image verifiable beyond doubt."

md



”
THE MASS OF
ARTIFICIALLY
CREATED IMAGES
LEADS TO A
LOSS OF TRUST,
INCLUDING
THE TRUST IN
REPUTABLE
MEDIA.“

Jonas Ricker

Quiz

WHICH PERSON IS REAL?

One of each pair of faces is real, the other one is artificially generated. Which faces are real?

The answers can be found on page 62.

All images are taken from the website whichfaceisreal.com.



1 a



1 b



2 a



2 b



3 a



3 b



4 a



4 b



5 a



5 b



6 a



6 b



EDITOR'S DEADLINE

The rabbits in the CASA Universe are startled: the seemingly well-secured access to Rabbit Mark's carrot stash has been hacked and all winter supplies have been stolen. The brave bunny Betty then starts looking for support at the nearby CASA Hub C – a mysterious place that is supposed to hold solutions for digital security. Thus begins the adventure of Betty the bunny, the protagonist of the first science comic from the Cluster of Excellence CASA. Along with Betty, the readers explore the Research Hub and learn about the research priorities and challenges that the scientists in the Research Hub C "Secure Systems" deal with on a daily basis. Find out how to read this and more CASA comics at no cost at:

➔ casa.rub.de/en/outreach/science-comics



Answers
DEEP FAKE-QUIZ
The following faces
are real:
1a, 2a, 3b, 4a, 5b, 6a

??

LEGAL NOTICE

PUBLISHER: Cluster of Excellence CASA at the Horst Görtz Institute for IT Security at Ruhr University Bochum in collaboration with the Corporate Communications Department at Ruhr University Bochum (Hubert Hundt, v.i.S.d.P.)

EDITORIAL ADDRESS: Corporate Communications Department, Editorial Office Rubin, Ruhr-Universität Bochum, 44780 Bochum, Germany, phone: +49 234 32 25228, rubin@rub.de, news.rub.de/rubin

EDITORIAL BOARD: Dr. Julia Weiler (jwe, editor-in-chief); Meike Drießen (md); Lisa Bischoff (lb)

PHOTOGRAPHER: Michael Schwettmann (ms), Dammstr. 6, 44892 Bochum, Tel.: +49 177 3443543, info@michaelschwettmann.de, www.michaelschwettmann.de

COVER: Sashkin – stock.adobe.com

PHOTOGRAPHS FOR TABLE OF CONTENTS: Michael Schwettmann

GRAPHIC DESIGN, ILLUSTRATION, LAYOUT:
Agentur für Markenkommunikation, Ruhr-Universität Bochum,
www.einrichtungen.rub.de/de/agentur-fuer-markenkommunikation

PRINTED BY: LD Medienhaus GmbH & Co. KG, Van-Delden-Str. 6-8, 48683 Ahaus, Germany, Tel.: +49 231 90592000, info@ld-medienhaus.de, www.ld-medienhaus.de

EDITION: 800

DISTRIBUTION: Rubin is published twice a year in German language; the regular issues are available from the Corporate Communications Department at Ruhr-Universität Bochum. The magazine can be subscribed to free of charge at news.rub.de/rubin/abo. The subscription can be cancelled by email to rubin@rub.de. The special issue 2021 is available from the Horst Görtz Institute for IT Security. In case of interest, please contact hgi-presse@rub.de.

ISSN: 0942-6639

Reprinting with reference to source and submission of proof copies