Can you trust what you see online?

A world leader gives a rousing speech, stirring up anger and resentment among their followers and opponents alike. A viral video shows a celebrity doing a hilarious dance routine. But is any of this real? Today, deepfake technology is so advanced that it is hard to know if you can trust what you see online. At the **University of Illinois Urbana-Champaign**, USA, **Dr Gang Wang** and **Jaron Mink** are investigating how we interact with artificially generated content and how to protect us from its harmful effects.

Dr Gang Wang

Assistant Professor

Jaron Mink

PhD Candidate

Department of Computer Science, University of Illinois Urbana-Champaign, USA

PROFILE

Field of research
Computer Science

Research project
Examining how people interact with phishing attacks and deepfakes, and protecting users from their harmful effects

Funder
US National Science Foundation (NSF)

TALK LIKE A ...

COMPUTER SCIENTIST

Artifact — in computer science, an error

Cybersecurity — the protection of computer systems and networks

Deepfake — any apparently authentic image, audio, video or text that is actually artificially generated

Machine learning — the development of computer systems to learn and adapt based on data, without following explicit instructions

Phishing — a type of social engineering attack in which attackers deceive users into revealing sensitive information or taking harmful actions

Social engineering — psychologically manipulating human users to compromise a computer system

URL (Uniform Resource Locator) — a link or website address

'Congratulations! You have won our top prize!'

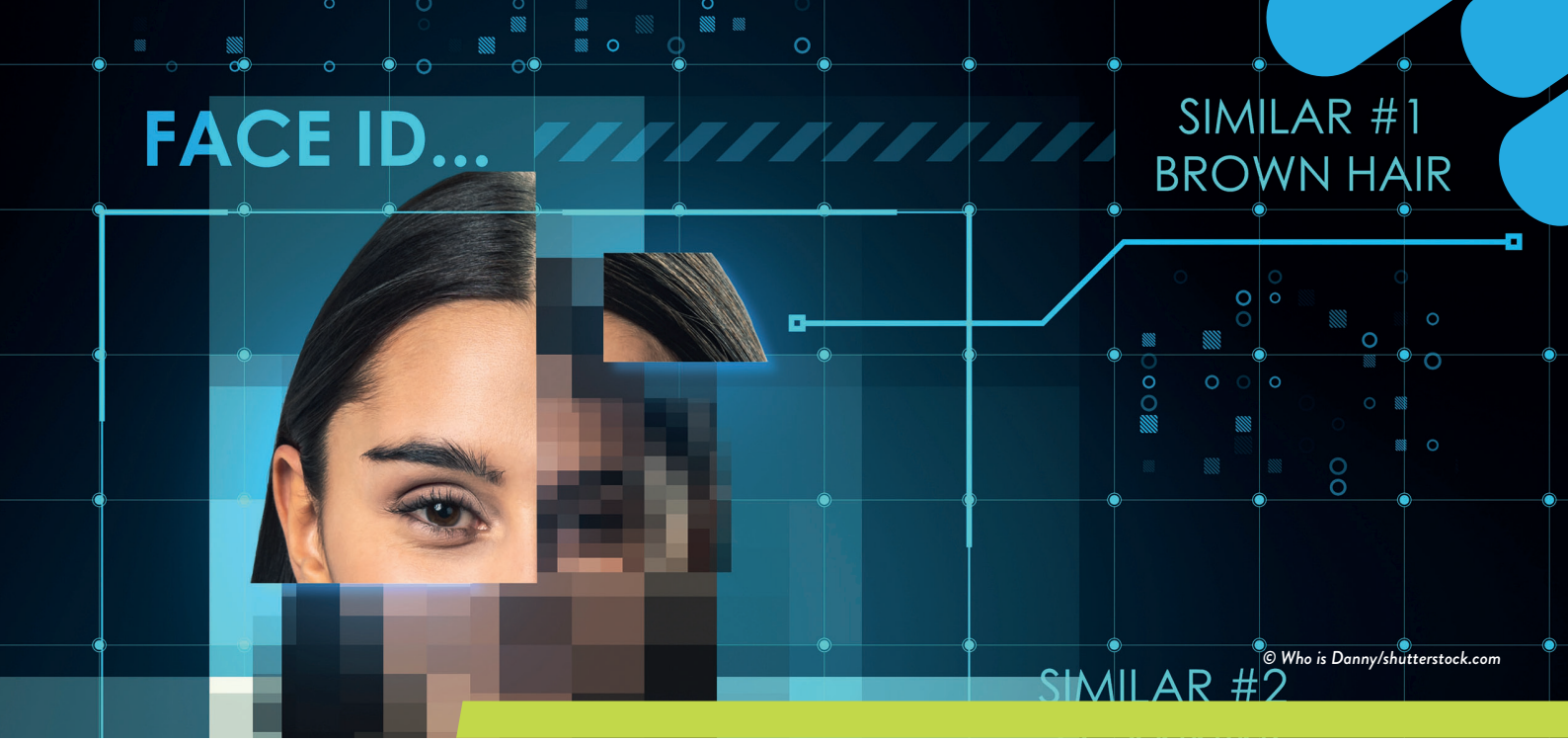
'Unusual activity has been detected on your TikTok account. Please click the link below and enter your login details to verify your identity.'

'We understand you have recently been in an accident that wasn't your fault. We can help you claim $10,000 in compensation!'

Have you ever received messages like these? They are examples of phishing, in which cybercriminals trick people into revealing sensitive information, such as passwords or bank details, or taking harmful action, such as installing malicious software. "In phishing attacks, attackers may gain the victim's trust by impersonating friends, co-workers, authorities or official sites," explains Dr Gang Wang, a computer scientist at the University of Illinois Urbana-Champaign. "These impersonations can occur through any communication channel, including emails, social media platforms, text messages and phone calls."

Deepfakes are another method that cybercriminals are increasingly using to trick people into believing information that is not true. "A deepfake is any content (e.g., video, image, audio or text) that appears to be authentic but has actually either been partially manipulated or fully generated by artificial intelligence," explains Jaron Mink, a PhD candidate working with Gang. While deepfakes can be used to create light-hearted entertainment (the internet is awash with fake videos of celebrities supposedly doing or saying funny or silly things), they can also be used maliciously to deliberately hurt people or spread false information.

"If deepfakes are being used for beneficial purposes, such as to enhance artistic projects by creating photorealistic scenes or for comedy entertainment, it should be made clear that viewers are interacting with artificially produced content," says Jaron. "Misuse arises when the technology is used to deceive others intentionally." Malicious deepfakes can harm the individuals who feature in them and they can threaten



national security. For example, in 2019, a deepfake social media profile successfully infiltrated Washington DC's political circle and connected with top government officials. It is, therefore, extremely important that deepfakes can be recognised and removed.

Gang and Jaron hope to not only improve cybersecurity by developing computer methods to detect phishing and deepfakes, but they also want to understand how we, as humans, interact with these phishing attempts and deepfake content.

How can we recognise phishing and deepfakes?

"There are signs that can help you identify phishing attempts," says Gang. These include requests for sensitive information (e.g., login details), urgent demands (e.g., 'Your account is about to expire. Update your details now.') and the presence of suspicious URLs. However, as artificial intelligence advances, it is becoming harder to distinguish deepfakes from authentic content. "There are still a few tell-tale signs," says Jaron. "Deepfake images of people may have distorted or asymmetric accessories or hands, while deepfake texts may contain repeated phrases or incoherent trains of thought."

How do computers protect us from phishing and deepfakes?

Does your spam inbox contain dodgy phishing emails as well as commercial advertising spam? Have you ever had a warning message when you tried to interact with an online account or website? If so, your computer has automatically detected these attack attempts and protected you from them. "Computers can be trained to detect phishing attacks by learning the differences between phishing attempts and genuine communications," explains Gang. "This can be done using machine learning algorithms, which statistically examine different features in the email or social media profile, such as keywords, URLs and images."

Unfortunately, it is still very difficult for computers to recognise deepfake content. "Both deepfake-generated content and the models used to detect them rely on very similar machine learning methods," explains Jaron. "This results in a cat-and-mouse game between cybercriminals generating new deepfake methods and cybersecurity experts developing new detection methods."

Therefore, while it is helpful for computers and social media platforms to detect phishing and deepfake content, computer scientists are trying to find more sustainable solutions. "We are developing machine learning explanation methods to help users recognise why messages and social media profiles may be suspicious," says Gang. Jaron believes we need to rethink how we view online content: "Perhaps we should focus on presenting content that we know is trustworthy, instead of focusing on removing content we think may be artificial."

How do people interact with deepfakes?

Gang and Jaron conducted a study to test people's responses to deepfake social media profiles, by creating fake profiles on the professional social network, LinkedIn, and asking whether people would accept a friend request from them. Some of these profiles contained artifacts, or small errors, in the artificially generated profile photo (e.g., a distorted background) or biography text (e.g., grammatical errors), or inconsistencies between the two (e.g., the biography states the person graduated in 1982 but the photo is of someone in their 20s).

Gang and Jaron discovered that while the presence of deepfake artifacts in a profile decreased the acceptance rate of friend requests compared to profiles without artifacts, 43% of participants accepted requests from the fake profiles they were presented with. "We asked participants why they accepted or rejected each friend request," says Jaron. "We discovered many people found it difficult to attribute certain artifacts to deepfake technology and not genuine human error." For example, incorrect grammar and inconsistent ideas in the biography text were commonly assumed to be due to poor communication and writing skills and so were not considered suspicious by participants.

What happens when people are warned about deepfakes?

Gang and Jaron also investigated whether people could be trained to detect deepfake artifacts. Of the 286 study participants, one third were not told that the LinkedIn profiles may contain deepfake content, one third were told, and one third were told and given in-depth training on how to recognise deepfake artifacts.

"When warned about deepfake content, we find significant decreases in friend request acceptances by participants," says Jaron. "However, even in such conditions, participants still accept friend requests from artifact-laden profiles." The deepfake recognition training did appear to help participants, as this group was better able to recognise poorly generated images as being artificial, while untrained participants were more likely to be confused, but unworried, by strange image artifacts.

However, Gang and Jaron discovered that warning people about fake profiles had some unintended negative consequences. "Trained participants may overcompensate in their search for deepfake artifacts, interpreting real profile characteristics as deepfake artifacts, and so decline genuine friend requests," explains Gang. Most worryingly, some wrongly-perceived artifacts stemmed from racial and gender stereotyping. "One participant became suspicious as they perceived that the name and image in a profile were stereotypically held by people of different gender and racial identities," says Jaron. "Encouraging users to try to differentiate real from fake profiles may instead cause them to differentiate between individuals that do and do not follow their held perceptions of what is 'normal'."

This highlights the need to ensure that deepfake training does not disproportionally harm marginalised and underrepresented communities. The future of cybersecurity, therefore, relies not only on computer scientists, but also social scientists, to create an inclusive and accessible internet.

About *cybersecurity*

As computer scientists, Gang and Jaron work in the field of cybersecurity, which involves protecting computer systems and the people who use them. They are especially interested in combatting social engineering and using machine learning to improve cybersecurity. Social engineering is the tactic of psychologically manipulating people into falling for cyber scams. "For example, social engineering may involve an attacker writing phishing messages in a way that creates a sense of urgency, fear or curiosity in the target user, luring them into giving away sensitive information," says Gang.

Cybersecurity experts are in a constant battle with cybercriminals, as new methods to attack and defend computers are continually being developed. As computer technology advances, there are many challenges for cybersecurity experts to solve. "These include improving security defences to handle increasingly complex and large-scale attacks, making security tools accessible for all users, and improving computers' abilities to respond to new and constantly changing attack tactics," says Gang. Machine learning can address some of these challenges. For example, it can enable computers to recognise patterns in large datasets and can automate key steps in the security defence process.

How can you stay safe online?

"Avoid over-sharing personal information (such as your name, address, phone number, birthday, etc.) and don't share sexually explicit images with anyone," advises Gang. "Such information may be used against you by abusers, for scams or harassment purposes." He also recommends verifying the privacy settings on all your social media accounts to ensure any personal information is only shared with your intended audience.

"Before engaging with information you find online, consider where it has come from," advises Jaron. If it is not from a well-known trustworthy source, he recommends remaining sceptical and not engaging with it. It is also important not to click on links with suspicious URLs, as these could take you to fake or malicious websites. Jaron advises taking care when connecting with people through social media. "Before accepting a friend request, make sure you've talked to that person in real life," he suggests. "And remember, even if the profile contains photos and details of a person you know, it might not actually be them sitting behind the computer communicating with you."

"It is very important to have a support system," finishes Gang. "Talk to trusted adults (such as parents, teachers or the police) if you believe you have encountered scams, phishing or other abusive behaviour online."

Pathway from school to *cybersecurity*

- At school, take any available computer-related courses. Mathematics is also important as computer science is very mathematical.
- You can learn computer coding and programming by exploring the many free websites and online tutorials available, such as Code Academy (**www.codecademy.com**). Once you know how to program, you can then learn about security and hacking (e.g., **www.pwn.college**).
- Jaron recommends joining your school or local computer club, and Gang recommends taking part in Capture the Flag (**www.ctftime.org/ctf-wtf**) events, where competitors solve cybersecurity challenges and attempt to (ethically) hack into systems to test cybersecurity defences.
- At university, degrees in cybersecurity, computer science, computer engineering or information technology will teach you the skills needed for a career in cybersecurity.
- "While in university, take courses that teach you about networking, operating systems, cryptography and security," says Gang. Explaining your work to others and educating the public about cybersecurity is also important, so it is a good idea to take classes in science communication and writing.

Explore careers in *cybersecurity*

- There is a wide range of career paths in cybersecurity. For example, security engineers build security infrastructure, security analysts monitor security systems, penetration testers (also known as ethical hackers) hack into computer systems to test security defences, and security consultants advise companies about cybersecurity.
- "Working in cybersecurity is intellectually stimulating and fun because the field is constantly evolving as attackers adopt new technologies and tactics," says Gang. "To defend against hackers, security professionals need to become better hackers themselves!"
- This article from Forbes explains how to get started on a career in cybersecurity: **www.forbes.com/advisor/education/how-to-get-into-cyber-security-field**
- Best Colleges has some useful information about degrees in the wider field of computer science and the careers these could lead to: **www.bestcolleges.com/computer-science/is-a-computer-science-degree-worth-it**

Meet *Gang*

As a teenager, I was primarily interested in computers because I enjoyed playing computer games. During college, my interests shifted to computer networks (the technology that enables the internet), then, during my PhD, I began to specialise in the security and privacy problems in networked systems.

As a student, I undertook internships at the Microsoft Research Lab, where I worked with researchers to tackle different security problems. I had the opportunity to interact with different production teams at Microsoft to understand the considerations and concerns from practitioners' perspectives. These experiences allowed me to test research ideas in practical scenarios. Personally, the internships helped me to better understand different roles in industrial research labs, which was very useful when it came to choosing a career path.

> ***AS A COMPUTER SCIENTIST, I ENJOY THE FREEDOM OF CHOOSING TO WORK ON EXCITING PROBLEMS.***

Computer science is a fast-moving field. Emerging technologies are changing almost every aspect of our lives, introducing new problems with respect to security and privacy.

As a computer scientist, I enjoy the freedom of choosing to work on exciting problems. I think can make a positive difference in the world. I also enjoy working with students. It is a rewarding experience to witness their growth and help them to become the next generation of computer engineers and scientists.

I am currently teaching my own kids how to program – computing skills are important life skills these days. And, in my free time, I still enjoy playing computer games!

Gang's *top tips*

1. Stay curious.
2. Learn how to program and start building things.
3. You will learn just as much by trying to break something as you did when building it. This is especially true for computer programming and security, so once you have programmed something, try to break it!

Meet *Jaron*

I enjoyed playing video games and board games when I was a teenager, as well as hanging out with friends. Don't discount how important goofing around with friends is to a healthy life!

My love for STEM came from a wonderful set of biology and physics high school teachers. These people were truly excited about teaching their subjects. While in their classes, it felt like what learning ought to be: curious, exciting and perspective-widening!

My high school didn't teach computer science, so I didn't realise that was my passion until later in life. In college, I spent the first year debating whether I wanted to study microbiology, chemistry or electrical engineering. By chance, I took a computer science course and immediately fell in love with the subject.

I love working in computer science because computers are embedded in nearly every scientific discipline and throughout society. With a skillset in computer science, you can improve algorithms for healthcare, discover novel ways to render graphics for video games, or work on theoretical mathematics about computing. Working with computers means being at the cutting edge of many problems the world is facing.

While most work in computer science is about building new systems, I love the human aspect of cybersecurity because it focuses on understanding the psychology of people, and how and why they act. Compared to theoretical work, human-centred computer science provides more immediate solutions to real-life problems.

For example, during an internship, I worked on developing an autism therapy app that allowed trained clinicians to remotely interact with parents of autistic children and provide treatment from afar. During the COVID-19 lockdowns, I worked with computer security researchers to ensure the university's contact tracing app was secure and protected the privacy of its users.

In my free time, I enjoy swing dancing, playing board games, hiking with friends and weight training. One day, I hope to learn to play jazz piano!

Jaron's *top tips*

1. Keep an open mind and don't just stay on a fixed path. It's extremely valuable to explore beyond that path.
2. Follow your interests. Pursue things because you're interested in them, not just because other people tell you to.
3. Make time for your friends, loved ones and hobbies. This is just as important as a successful career as they will support you and bring you joy on your journey through life.