KPMG | Reality Defender®

Deepfakes: Real threat

As artificial intelligence grows ever more powerful and sophisticated, it has become easier to create fake content than detect it.

kpmg.com

Introduction

Social engineering—an act of deception aimed at compromising personal or sensitive information—remains an effective tactic for bad actors bent on creating disruption or illegitimate financial gain. More and more, these pernicious campaigns, which include phishing and spear phishing, employ fraudulent deepfake content and are becoming more brazen by the day.

Over the last several years, the perpetrators behind these schemes are looking for bigger game than individual consumers or public figures. Creative, ambitious cyber criminals with access to the latest technology have started to focus on more profitable targets—corporations, institutions, and sovereigns—many of which are ill prepared to defend against this threat.

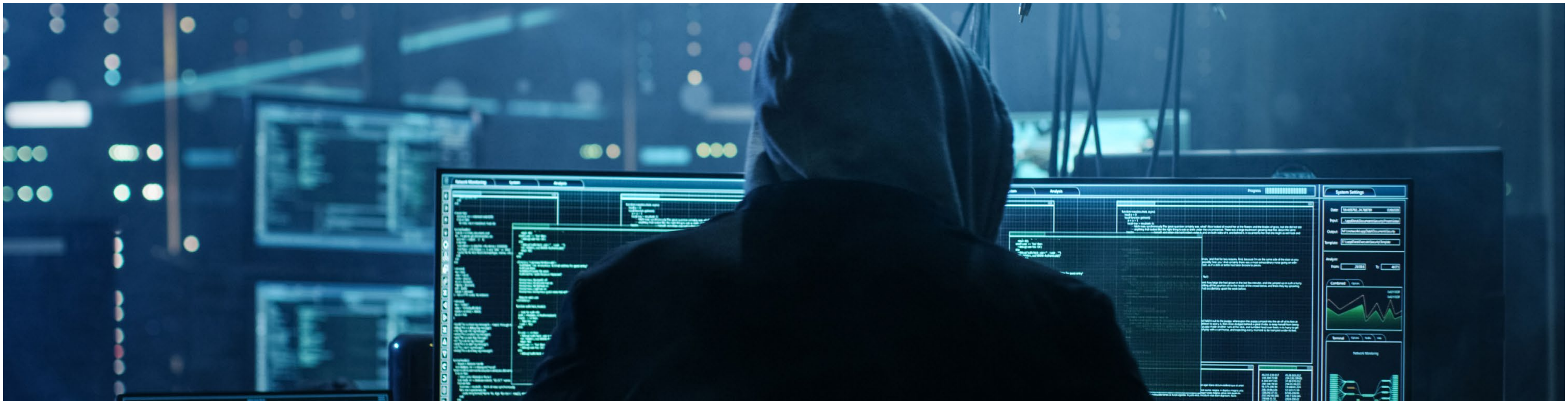
In a Private Industry Notification in March 2021, the FBI revealed its expectation that "malicious actors almost certainly will leverage synthetic content for cyber and foreign influence operations in the next 12-18 months." The notice states further that the FBI believes these actors will use AI techniques broadly, primarily to facilitate existing social engineering schemes.[1]

The consequences of cyberattacks that utilize synthetic content—digital content that has been generated or manipulated for nefarious purposes—can be vast, costly, and have a variety of socioeconomic impacts, including financial, reputational, service, and geopolitical, among others.

In this paper we describe the threat deepfakes pose, the challenges companies are encountering detecting this content, the need for organizations to take the risk seriously, and strategies security leaders can consider for their networks.

[1] Federal Bureau of Investigation, Private Industry Notification, March 10, 2021

What is a deepfake?



Simply stated, deepfakes are synthetic media files. They are "synthetic" in that existing imagery, video, or audio—typically featuring a specific individual—is manipulated and replaced with another person's face or voice. This work is done using generative artificial intelligence (AI)-powered neural networks, also known as Generative Adversarial Networks (GANs), that process information, create patterns, and learn much like the human brain does.

Today, widespread availability of sophisticated computing technology and the growing accessibility of AI enables virtually anyone to create highly realistic fake content. In fact, the number of deepfake videos available online is increasing by 900 percent annually.[2] The manipulation of content to influence audiences is not new, but the line between what's real and what's fake has become razor thin. Through disruption and deception, deepfakes can cause significant damage both to individuals and institutions.

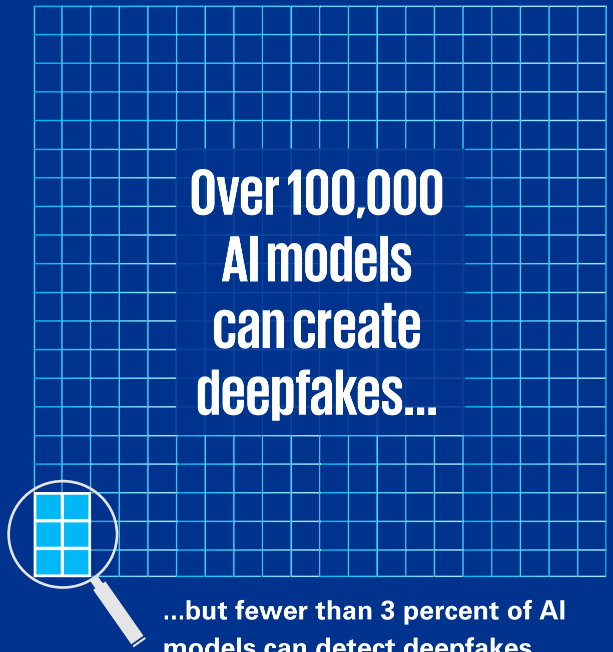
According to Dr. Nasir Memon, professor of Computer Science and Engineering at the New York University Tandon School of Engineering, the prevalence of deepfakes is growing, including the development of "Deepfake-as-a-Service" opportunities on the dark web. There are many other monetization opportunities for cyber criminals, with the potential of misinformation being weaponized by rogue nation-states expanding as threat actors add deepfakes to their attack arsenal.[3]

[2] Reality Defender, 2023

[3] Nasir Memon, "Deepfakes, shallowfakes & cheapfakes—Seeing is believing," eniineering.nyu.edu, March 28, 2021

Difficulty of identifying deepfakes

GANs are capable of creating media that is virtually undetectable to cybersecurity experts. There are more than 100,000 models that have been developed to create and/or detect deepfakes, of which only about 3,000 are able to identify deepfakes properly. Most models are only able identify the "cheapfakes," which are unsophisticated copies generated through the most simplistic manipulations.



Over 100,000 AI models can create deepfakes...

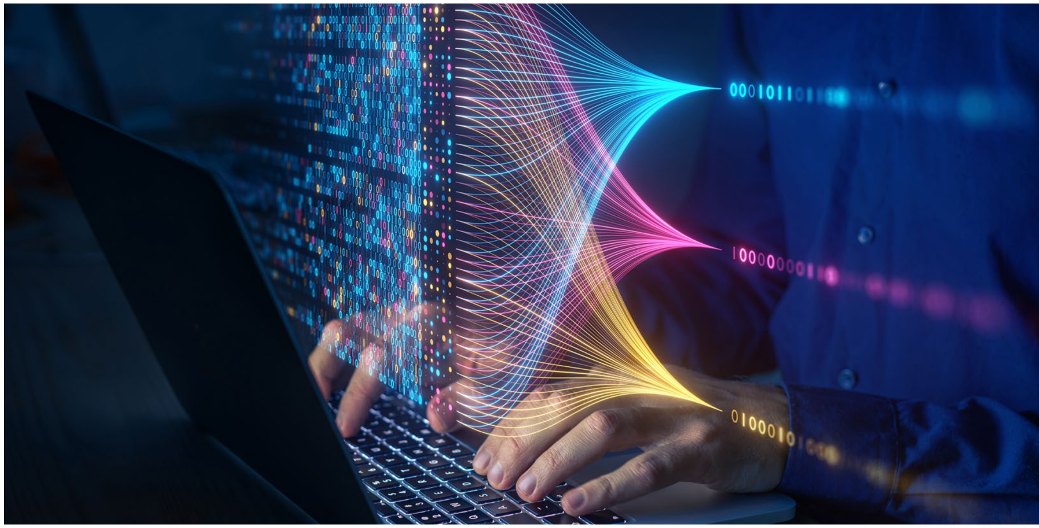
...but fewer than 3 percent of AI models can detect deepfakes

Source: Reality Defender, 2023

Generative AI connection

Whether it's cheap fakes or deepfakes, the economic model has shifted significantly because of Generative AI models. Recent technological advancements make it possible for anyone with a computer to create deepfakes—and open-source code, free mobile applications, online tutorials, and inexpensive AI service providers have greatly expanded access to the necessary tools.[4]

As a result, criminals are seeing a great leap forward in their ability to commit fraud, extort large sums of money, damage reputations, and even disrupt national security. AI is now being used to alter maps, imagery, and X-rays, generate text, and even create realistic artwork. And AI-generated digital avatars are increasingly able to hold conversations, exhibit emotion, and make realistic human gestures.[5] The evolution of manipulated content has come a long way in a short period of time. Deepfakes are close to being readily available for a variety of purposes, not all of which are well intended.



[4] [5] Ali, O., et al. "Trends in AI from Red and Blue Team Perspectives: Synthetic Data in a Data-Driven Society vs Sentiment Analysis, NATO Strategic Communications Centre of Excellence," December 2022

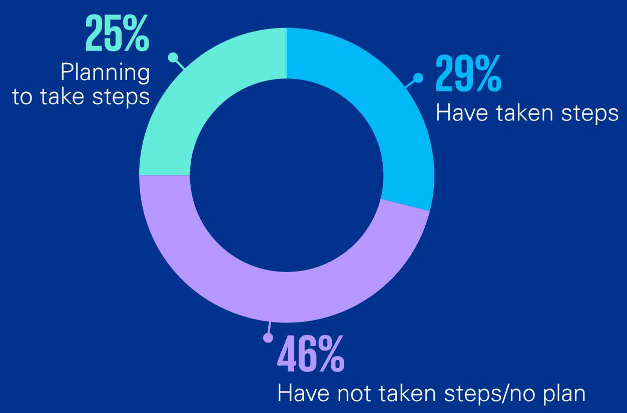
From parlor trick to enterprise risk

As a risk factor, deepfake content is not merely a concern for social media, dating sites, and the entertainment industry—it is now a boardroom issue. Case in point, nearly all respondents (92 percent) of a recent KPMG generative AI survey of 300 executives across multiple industries and geographies say their concerns about the risks of implementing generative AI is moderately to highly significant. The survey also found the three most prevalent management and mitigation focus areas for these business leaders are cyber security (53 percent), privacy concerns around personal data (53 percent), and liability (46 percent).[6]

And in the aforementioned Private Industry Notification, the FBI suggests synthetic content may be used in a method of attack the bureau has dubbed Business Identity Compromise (BIC), which will leverage advanced content generation and manipulation techniques to create synthetic personas based on existing employees. This emerging vector could have significant financial and reputational impacts to businesses and organizations.[7]

Furthermore, a 2021 survey of professionals at firms across more than 13 industries, including IT/data services, healthcare, financial services, and manufacturing, found that more than 80 percent of respondents said deepfakes pose a potential risk to their business, but 29 percent say they have taken steps to protect themselves. Nearly half (46 percent) said their organization had no plan to mitigate the threat.[8] Then, in 2022, a similar survey of insurance professionals revealed that more than 80 percent of respondents in that sector are concerned about manipulated digital media, as well. While the vast majority of insurance respondents noted concern about deepfakes, only 20 percent reported taking any action.[9]

Companies are slow to act*



Since social engineering attacks, such as phishing and spear phishing, typically rely on some form of impersonation, deepfakes are the perfect addition to the corporate cybercriminal's tool bag.[10]

Unfortunately, many corporate decision makers have a very limited understanding of deepfakes and the threat they pose. Most leaders don't yet recognize that manipulated content is already a problem—rather than an emerging or an eventual one—that affects virtually every industry. Even more troubling, many don't yet believe the business risk is significant.

Bottom line, the deepfakes threat and accompanying financial, reputational, and service implications are very real for companies and institutions. And with the technology improving at a breathtaking pace, deepfake-related concerns are scaling more rapidly than they did with phishing 25 years ago.

* Chart: Attestiv, Deepfakes: The Business Threat, Inaugural Deepfakes in Business Survey, 2021

[6] KPMG Generative AI survey, 2023

[7] Federal Bureau of Investigation, Private Industry Notification, March 10, 2021

[8] "Inaugural Deepfakes in Business Report," Attestiv, Inc., May 2021

[9] "Deepfakes: A Threat to Insurance?" Attestiv, Inc., May 2022

[10] "The Deepfake Problem," student thesis, University of Twent The Netherlands, August 2022

Cybersecurity and mitigation approach

As the pace of technology innovation quickens, the opportunities from a detection and defense perspective are expanding. Unfortunately, would-be attackers are exploiting these advances, as well. Many businesses have ready access to today's deepfake toolkit—GANs, autoencoders, and generative and discriminative algorithms—but if they face increased hurdles to implementation because of external regulations and internal bureaucracy, the bad actor side of the equation may be positioned, for the moment at least, to take advantage of the opportunity first.

Technology: AI

Security opportunity
Improved proactivity for mitigating cyber events through intelligent threat and anomaly detection, as well as better response times

Attacker risk
Ability to successfully optimize social engineering strategies — in particular, phishing and spear phishing — to rapidly improve attack efficacy

Source: KPMG Ignition, 2023

Attacks involving deepfakes are a logical component of social engineering. For cyber security purposes, they should be viewed as related, because the malicious potential of deepfakes intensifies when they are used as the focal point of social engineering schemes.

Currently, deepfakes in isolation are a small component of the overall suite of cybersecurity concerns, but because the incidence of attacks featuring manipulated content is growing, chief information security officers (CISOs) and their teams are taking notice. In fact, according to a 2022 survey of 125 cybersecurity and incident response professionals, 66 percent said they had experienced a security incident involving deepfake use in the past 12 months, an increase of 13 percent year-over-year.[11]

Mitigation strategy and associated costs

From a mitigation perspective, the major concern with respect to advanced deepfake technologies is the funding required for detection, from maintaining the appropriate computing power, forensic algorithms, and audit processes to the talent needed to employ these tools. CISOs are encouraged to initiate conversations with senior decision makers to ensure budgets match the threat and keep technology up to date by ensuring software updates are installed as soon as they are released.

Similarly, organizations should implement zero-trust and multi-factor authentication processes to minimize the risk of these cyber resources being used to compromise legitimate network users.

With identity at its core, zero trust enables organizations to evaluate whether a user is properly authenticated; isolate the resource the user is attempting to access; determine if the request is from a trusted, stolen, or third-party device; and confidently decide whether access should or should not be granted.

The emergence of zero trust represents a mindset shift in which CISOs and their teams assume compromise in connection with system access, and makes security decisions on the basis of identity, device, data, and context.

The average cost of a cyber security breach was $1.76 million less for organizations with a mature zero-trust methodology relative to those who don't employ zero-trust.[12]

[11] VMWare Global Incident Response Threat Report, 2022

[12] "Cost of a Data Breach Report," Ponemon Institute/IBM Security, 2021

Why is the deepfake situation getting worse, not better?

Today's pace of innovation has spurred rapid transformation, incentivizing fast movement across the enterprise, but too often at the expense of security. Businesses are innovating, but the room for error is growing. This digital acceleration has expanded the attack surface and greatly increased the number of assets in need of advanced security, ultimately putting organizations at risk. At the moment, authentication capabilities are largely insufficient to deal with the steady advance of deepfake technology.

Being an innovation leader is critical, but security needs to be brought along on the transformation journey. The opportunity for cyber incidents continues to rise with new technology innovation, a more active, sophisticated, and political threat landscape, and a greater reliance on data. Today's rapid pace of innovation often results in security being left behind in business transformation efforts: 39 percent of global IT leaders do not believe that their organization's security measures have kept pace with their digital transformation initiatives.[13]

Remote work poses a deepfake concern

Many security leaders believe that home-based employees are particularly vulnerable to manipulated content attacks, creating consternation as more workers leave the office behind. Perhaps even more troubling, the FBI has warned that deepfakes are being used by bad actors to apply for a broad array of remote IT and programming positions in an effort to gain access to personal, financial, and other proprietary information.[14] These schemes can have significant financial repercussions. Indeed, the average impact of a data breach was more than $1 million higher when a remote working arrangement was a factor in the breach.[15]

To make matters worse, hybrid working has expanded the attack surface, raising the number of potentially vulnerable endpoints. Adding to the challenges, shadow IT within organizations—information technology activities that occur outside the purview of a firm's actual IT group—often includes third-party business applications and software as a service use over which CISOs and CIOs have limited visibility or understanding of the possible exposures.

[13] IT Leader Survey, Veritas Technologies, 2021

[14] Federal Bureau of Investigation, Private Service Announcement, June 28, 2022

[15] "Cost of a Data Breach Report," Ponemon Institute/IBM Security, 2021

The human variable

Humans are inherently unpredictable and difficult to control. As phishing and other social engineering attacks become more sophisticated, the pressure to mitigate human nature grows ever more important. Similarly, employee, vendor, and customer responsibility around cybersecurity is increasing. Beyond annual training modules, cyber security in the deepfakes age demands persistent personal engagement that draws on observational learning and behavior-reinforcement techniques to create cohesion on the need for secure behavior.

By investing in the human element of cyber security, organizations can foster workforces that are not only savvier about and aware of cyber security but also a crucial extension of the cyber security team through their commitment to keeping the organization safe. Internal security controls should be easy to use, or employees may be motivated to bypass these processes. Indeed, a recent study found that 54 percent of employees fell for phishing scams over the preceding 12 months because the email looked legitimate, and 52 percent said the attack was successful because they believed it had come from a senior executive.[16]

A holistic approach to protect an organization requires an investment in people to ensure that employees both understand the tenets of cyber security and embrace their role in supporting security efforts by making secure behaviors an integral part of their daily lives.[17] (See page 14 for a list of tangible steps CISOs and their teams should consider as they seek to counter the deepfakes threat.)

Compliance is a full-time job, but it isn't enough

Regulatory structure changes so quickly that companies spend resources just focusing on compliance, rather than investing in longer-term solutions. Current regulations/laws in the U.S. do not have any actual relevance in tackling the issue of deepfakes. No particular laws exist.

This will soon change, and companies need to be ready for a regulatory shift in connection with AI-related matters. In May, the CEO of OpenAI—creator of the much-discussed ChatGPT—testified before a Senate subcommittee, mostly agreeing with senators for the need to regulate AI and proposing the creation of an agency that would issue licenses for the development of AI models.[18]

And in China, regulations to scrutinize and protect individuals from deepfakes took effect in January.[19]

[16] Jeff Hancock, "Psychology of Human Error," Tessian, 2022

[17] KPMG, "Human firewalling: Overcoming the human risk factor in cyber security," 2021

[18] Cecilia King, "OpenAI's Sam Altman Urges AI Regulation in Senate Hearing," The New York Times, May 16, 2023

[19] Brenda Goh, "China's rules for 'deepfakes' to take effect from Jan. 10," Reuters, December 12, 2022

Deepfakes in the real-world

Fraud on social media platforms appears in myriad ways. Deception in individual accounts often starts when unorganized or loosely organized bad actors create multiple "sock puppet" accounts—a profile that shields the user's true identity for the purpose of artificially elevating a keyword or topic. The user posts misleading or false messages under the guise of a real person who, in actuality, does not exist.

These platforms also experience concentrated organized efforts from teams of bad actors—sometimes working on behalf of a nation-state or political entity. These groups can activate hundreds of thousands, if not millions, of "bot" accounts to spread disinformation and otherwise harmful content. These automated accounts often work together to give them the appearance of legitimacy, "liking" and following each other as if they were controlled by real people.

While the messaging in the posts from these accounts often is created by humans, the images used to legitimize the fake account often are visual deepfakes, AI-generated photos or videos created to deceive. This visual content can be created using a number of models and tools, from off-the-shelf and open-source GANs to more advanced proprietary methods.

Today there are hundreds of ways for fraudsters to produce visual deepfakes, with new and more advanced models being introduced seemingly every week to support the growth of a nascent Deepfakes-as-a-Service business model.

Anatomy of a call center voice fraud scheme

Most people can detect low-quality deepfakes on sight and sound alone. Visual glitches and obvious or even limited facial twitches or expressions are often signs that lead the viewer to discern the difference between a real speaker and an AI-generated "persona." Speech may not sync with person speaking in a video, or the voice may sound unnatural or robotic.

In contrast, higher quality deepfakes are a challenge for even the most perceptive viewer/listener. A video or image generated using a sophisticated deepfake method often can only be identified by equally advanced deepfake detection platforms, such as those developed by several prominent public and private software companies and a number of leading universities. In the absence of such detection platforms, this content leaves viewers vulnerable to deceptive messages, particularly political propaganda, disseminated by fictitious entities.

These systems model facial motion, the human vocal tract, and biometric voice signatures to determine if the sample is biologically plausible or likely fake. They can also watermark generated text, and flag word choices and sentence length that don't appear to have been generated by a human.[20]

In the current environment, creating an advanced and difficult-to-detect audio deepfake is as simple as using off-the-shelf hardware and readily available generative AI models, often in real-time.

Like other deepfakes, the methods used to create audio deepfakes are perpetually advancing, often with dire consequences.

For example, in January 2020 the manager of a Japanese bank based in Hong Kong took a phone call from who he thought was the director of the company. The manager was told the company was preparing for an acquisition needed $35 million to be transferred to a bank in the United Arab Emirates, which the manager authorized. Seemingly, there was no reason to doubt the call was legitimate. An investigation subsequently found that deepfake voice technology was used to imitate the director.[21]

Obtain audio file

Bad actors obtain a recording of their target's voice and break it down into syllables and sounds.

Online tools to replicate a target's voice

The audio is reassembled to create a new voice message that sounds identical to the speaker.

Fraudsters use fake voice with call centers

Bad actor engages call centers to execute financial fraud, password resets or account takeover.

[20] Matthew Hutson, IEEE Spectrum, "Detection Stays One Step Ahead of Deepfakes—for Now: The spread of AI-generated content is keeping the tech designed to spot it on its toes," March 6, 2023

[21] Thomas Brewster, Forbes, "Fraudsters Cloned Company Director's Voice In $35 Million Heist, Police Find," October 14, 2021

Take action: What companies can do now

Deepfakes appear poised to destabilize organizations financially and operationally. To address the growing access and identity risks presented by synthetic and manipulated content, as well as respond to an expanding regulatory environment, enterprises are encouraged to consider new processes, tools, and strategies to better secure their systems, data, and infrastructure.

In a cloud-based, zero-trust future, users likely will no longer need to be “on network” through persistent virtual private network connections. Conditional access will come from the trust and assurance that is engendered by the devices people use, and the authentication and decisioning processes organizations implement. In this new reality, security professionals need to be vigilant in regard to fraudsters and their attempts to use deepfakes to override biometric and other identity verification solutions.

Develop a robust cyber security culture and hygiene

Throughout business activities and decision making—including work-from-home arrangements—prioritize a robust cybersecurity culture that engages employees and highlights the behaviors they need to understand to create their own cybersecurity hygiene, and function as human firewalls.

Strengthen identity confirmation to counter sophisticated fakes

Eliminate implicit trust in network systems and continuously validate throughout each digital interaction. Implement a zero-trust model, multi-factor authentication, behavioral biometrics, single sign-on, password management, and privileged access management.

Improve organizational risk intelligence

Identify key risk indicators and quantify the financial impact of security risks and threats posed by manipulated content.

Secure by design

Mitigate the human variable through secure by design by embedding security throughout product lifecycles, connecting cyber strategy to the entirety of the business.

Invest in technology for better prediction, detection, and response

Make technology a budgetary priority to enable new and innovative approaches to cyber security problem solving.

A broad-based threat

Deepfakes endanger finances, security, reputation, brand, and even sovereignty throughout the enterprise, consumer, and government ecosystems.

<table>
<tr><th colspan="3">Enterprise</th><th colspan="5">Consumer</th><th colspan="3">Government</th></tr>
<tr><th>Financial Services</th><th>Insurance</th><th>Health Services</th><th>Marketplace + Auctions</th><th>Social Media</th><th>Adult</th><th>Dating Platforms</th><th>News</th><th>Political</th><th>Government</th><th>Defense</th></tr>
<tr><td colspan="8">User Onboarding Fraud (KYC)</td><td></td><td></td><td></td></tr>
<tr><td colspan="11">Content Moderation and Analysis</td></tr>
<tr><td colspan="11">Voice Verification Fraud</td></tr>
<tr><td colspan="11">Email + Messaging Threats</td></tr>
<tr><td colspan="11">Fraudulent Internet Posts (Companies, Products, People Figures)</td></tr>
<tr><td colspan="8">Age Verification Fraud</td><td></td><td></td><td></td></tr>
<tr><td colspan="4">Financial & Legal Document Fraud</td><td></td><td></td><td></td><td colspan="4">Geospatial Mapping Manipulation</td></tr>
</table>

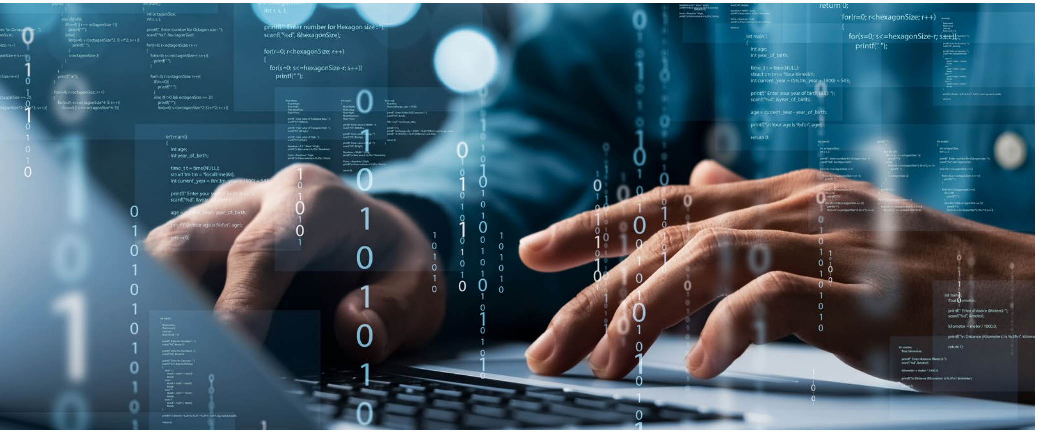
Source: Reality Defender, 2023

Conclusion

Awareness and adaptability are key

The wide-ranging promise of AI is truly exciting. However, the technology underlying these advances is now also available to those who would use it for mischief… and far worse. That's the blessing and the curse: generative AI models can learn to do remarkably precise and reliable work designed to illuminate or deceive. From an enterprise perspective, the questions in relation to AI, ML, and deep learning are around augmenting existing systems to detect and respond to these threats and malicious software attacks.

In general, security organizations have lagged in their ability to predict, detect, and respond to these threats in an automated manner. This is where the power of deep learning systems can be leveraged to ensure that cyber security systems have the ability to continually learn and improve their defense mechanisms. The key for security professionals is understanding—and keeping up with—the pace and speed at which cybercriminals use these tools.



How KPMG can help

The smartest businesses don't just manage cyber risk, they use it as a source of growth and competitive edge. Technology makes many things possible, but what's possible isn't always safe. Your cyber security must build resilience and trust as cyber threats grow in volume and sophistication, and as technology becomes essential for meeting the needs of your customers, employees, suppliers, and society.

We can help you create a resilient and trusted digital world—even in the face of evolving threats. Our professionals bring a combination of technological expertise, deep business knowledge, and creativity along with passion to protect and build your business.

KPMG has experience across the continuum—from the boardroom to the data center. In addition to assessing your cyber security and aligning it to your business priorities, we can help you develop advanced approaches, implement them, monitor ongoing risks, and help you respond effectively to cyber incidents. So, no matter where you are on the cyber security journey, KPMG can help you reach the destination.

Strategy and governance: Turn risk to competitive advantage.

Cyber transformation: Accelerate your initiatives in an agile world.

Cyber defense: Confidently seize opportunities.

Cyber response: Operate with confidence in a digital world.

About the author

Matthew Miller
Principal
Cyber Security Services
KPMG
matthew.miller@kpmg.com
212-954-4648

Matt is a Principal in the New York office of KPMG LLP and leads U.S. Cyber Security Services Financial Services. Matt established and leads KPMG's Center for Cyber Analytics Research focused on establishing next generation analytics, developing AI and ML cybersecurity solutions and securing AI models. Matt has more than 24 years of cyber experience insider threat and internal fraud, threat intelligence, vulnerability analysis, 3rd party risk, quantitative and qualitative risk assessment, and incident management. In addition to managing programs or advising clients, Matt has published and presented on many subjects, including leveraging capability maturity models to improve risk management, addressing vulnerability in technologies and critical business applications, and establishing governance and metrics to enable effective risk management programs.

This research was a collaboration between KPMG and Reality Defender, a deepfake detection platform used by enterprises to flag fraudulent identities, transactions, and online media. The firm's API and web app provide realtime risk scoring, email alerts, and forensics review capabilities. We thank Ben Colman, Co-Founder & CEO of Reality Defender, for his contributions and insights in this paper.

Reality Defender®

Related thought leadership:

KPMG

Using generative AI to strengthen cyber security

kpmg.com

KPMG

How synthetic identity fraud evades detection

KPMG

The Perfect Cyber Crime: Synthetic Identities and COVID-19 Relief Funds

KPMG

Detecting the undetectable: Strategies for identifying synthetic identity fraud

KPMG

Synthetic identity fraud: A $6 billion problem

Some or all of the services described herein may not be permissible for KPMG audit clients and their affiliates or related entities.

The information contained herein is of a general nature and is not intended to address the circumstances of any particular individual or entity. Although we endeavor to provide accurate and timely information, there can be no guarantee that such information is accurate as of the date it is received or that it will continue to be accurate in the future. No one should act upon such information without appropriate professional advice after a thorough examination of the particular situation.

DASD-2023-13188

kpmg.com/socialmedia