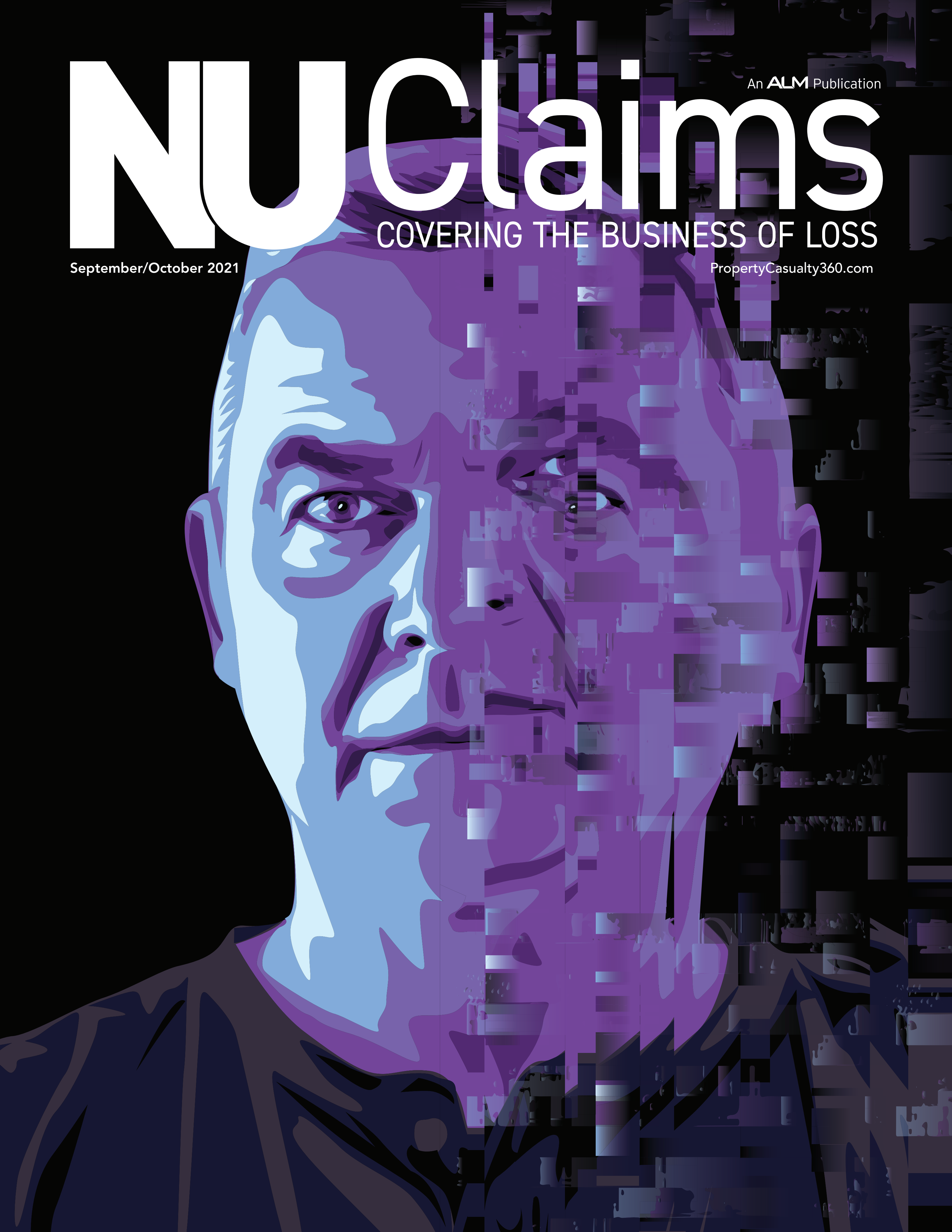
NU Claims

An ALM Publication

COVERING THE BUSINESS OF LOSS

September/October 2021

PropertyCasualty360.com



DEEPFAKES: AN INSURANCE INDUSTRY THREAT

By Nicos Vekiarides | Illustration by Daniel Hertzberg

IF YOU ARE FAMILIAR WITH PHOTO AND VIDEO EDITING tools, then you have probably heard of deepfakes, an emerging breed of artificial intelligence (AI) enhanced videos that have demonstrated the ability to blur reality in ways that are extremely difficult for humans or even machines to detect. Unlike conventional video editing, deepfakes utilize AI to alter or synthetically generate videos, bringing a new level of realism, without the forensic traces present in edited digital media. While these advanced fakes may sound like science fiction, many researchers have concluded it is only a matter of time before deepfakes become nearly undetectable to the human eye and subsequently undetectable even to elaborate forensic tools.

Whereas deepfakes have largely manifested as a novelty on social media, deepfakes and similar AI-generated photos and videos can pose a significant threat to industries that make important financial decisions on the contents of photos and videos, such as insurance. The ability to distort reality, in ways that are difficult or impossible to detect, significantly increases the risk of digital media fraud in insurance claims at a time when many carriers have rapidly adopted self-service as a way to process claims during the COVID pandemic.

Earlier this year, the FBI sounded the alarm that deepfakes are a new cyberattack threat targeting businesses. As a result many organizations are pondering strategies to mitigate the risks and potentially undesirable outcomes that may result.

> DEEPFAKES OR SYNTHETIC MEDIA CAN EFFECTIVELY BE USED TO FILE FRAUDULENT CLAIMS, CREATE FRAUDULENT INSPECTION REPORTS, AND EVEN ESTABLISH THE EXISTENCE AND CONDITION OF ASSETS THAT DO NOT EXIST.

Deepfake awareness

To help promote awareness of the danger deepfakes pose in the corporate realm, Attestiv recently surveyed U.S.-based business professionals about the threats to their businesses related to synthetic or manipulated digital media. The survey also inquired about their plan of action and defense strategies.

Not surprisingly, over 80% of respondents acknowledged that deepfakes posed a threat to their organization. The top three deepfake concerns included:

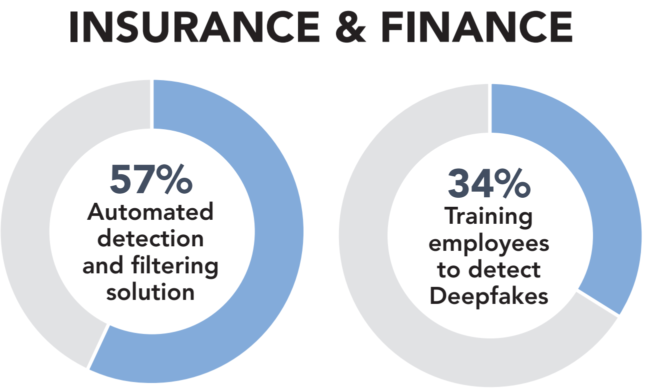
1. Reputational threats
2. IT threats
3. Fraud threats

While every cyberthreat poses reputational and IT risks, the fraud aspect is most relevant for the insurance industry, as it relies on digital photos, videos and documents to make business decisions and is already subject to tens of billions of dollars in annual fraud in the U.S. alone.

On the question of what steps organizations will take to protect themselves against altered digital media, less than 30% of respondents revealed having any defense strategy in place. While the amount of inaction exposes a problem, one consolation is that another 25% of respondents said they are planning to take action, meaning they recognize the threat and a solution is in the works. On the other hand, that leaves a total of 46% of respondents without a plan or knowledge of the plan.

Perhaps ironically, the results were slightly worse for insurance, where only 39% of respondents indicated they are either taking or planning steps to mitigate the risk of deepfakes. These numbers were surprisingly lower than the mean, given other industries might be less susceptible to digital media fraud.

When asked "What's the best defense organizations can take against altered digital media?" the results showed over 57% of respondents in both insurance and finance sectors felt the best defense was an automated detection and filtering solution, while 34% felt training employees to detect deepfakes was the best solution. This result proved both encouraging and somewhat distressing.



Automated detection and filtering solutions are indeed a viable approach to stopping deepfakes, as there are currently solutions on the market employing technologies such as blockchain or AI to prevent or detect manipulated media. On the other hand, training employees to detect deepfakes is a far from viable solution given the likelihood that they are rapidly becoming undetectable to human inspection. For some companies, there may be a need for further education regarding the deepfake threat and the trajectory the technology is taking.

Help from industry standards

Back in September of 2019, Facebook partnered with other companies and academia to launch the Deepfake Detection Challenge, in the hope of getting ahead of the substantial disinformation threat deepfakes pose on social media. Many entrants built technologies to detect deepfakes and manipulated media, and the results published in June 2020 were promising, albeit less than stellar. The top performer registering an accuracy of 65% on a black box data set. While this was a good start, it left a lot of room for improvement.

Around the same time, other working groups launched such as the Content Authenticity Initiative (CAI), a cross-industry initiative that allows for better evaluation of content provenance, started by Adobe, in partnership with Twitter and the New York Times. Similarly, the C2PA was founded in February 2021 by Microsoft and Adobe to deliver technical standards for content provenance and authenticity.

While standards have started the march toward helping thwart deepfakes across various industries, insurance companies have the choice of waiting or developing an interim plan.

An approach to protecting insurance

You may have seen the website “This X Does Not Exist,” a clever website using generative adversarial networks, the technology behind deepfakes, to create synthetic people, vehicles, cats, rental homes and the like. While it’s an entertaining diversion, it’s by no means a stretch to apply the same technology to create fake accidents or home damage that can exaggerate or create a fake insurance claim. What’s more, now that this proverbial cat has been let out of the bag, it is not poised to disappear any time soon.

So what should an insurance carrier do? Bringing some form of automated deepfake protection to insurance is the most viable solution for protecting against this new breed of fraud. But, how can it be implemented into existing processes for filing claims?

As it turns out, some of the processes may not need to change at all. For instance, any claims photos that are gathered by adjusters are already going through a trusted third party. While no inside or outside party is immune to fraudulent behavior, a trusted stakeholder would likely be risking their job and reputation by filing false claims. Simply put, the cost of committing claims fraud would be very high.

On the other hand, any processes driven by the insured in a self-service manner are susceptible to manipulated or fake media. Consider:

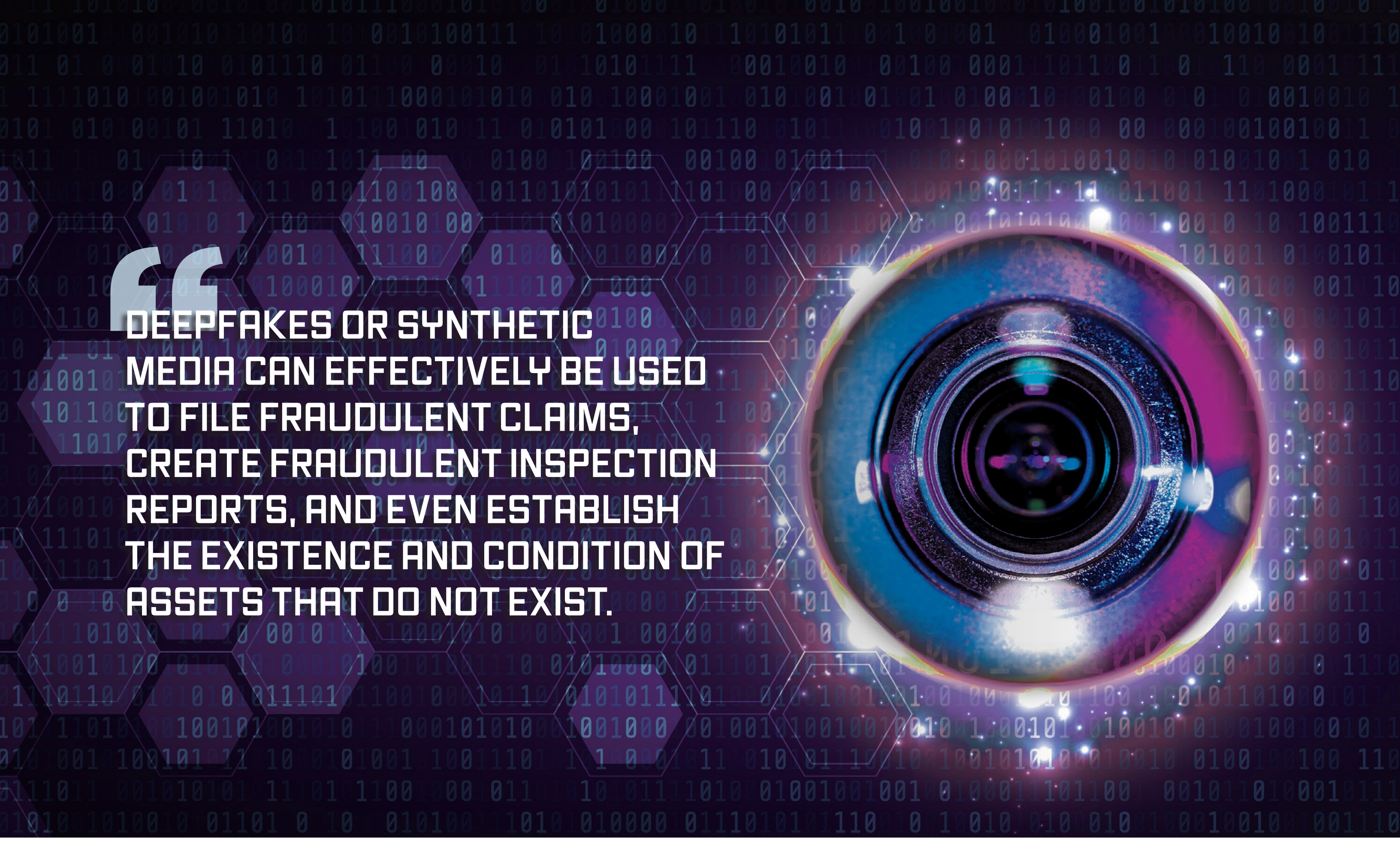
- Auto or home claims
- Inspections for underwriting or purposes of loss control
- Establishing existence and condition of assets during underwriting

Deepfakes or synthetic media can effectively be used to file fraudulent claims, create fraudulent inspection reports, and even establish the existence and condition of assets that do not exist. Think claims for exaggerated damage from a nearby hurricane or tornado or claims for items that don’t even exist; i.e., a non-existent Rolex watch that got insured and mysteriously went missing.

Does this suggest going back to using human adjusters and inspectors for important claims? While taking a step backward to manual inspection might help to eliminate the deepfake threat, a layer of protection against the deepfakes in self-service processes would serve better without undoing years of digital transformation. Moreover, with many claims processes moving to straight-through processing, with no human intervention required aside from exceptional cases, two in-line approaches are suggested for implementing a layer of defense:

1. **In-line Detection:** Using AI and rules-based models to detect deepfakes in all digital media submitted. Similar to the Deepfake Detection Challenge mentioned earlier, apply AI-based forensic analysis to every photo or video prior to processing a claim.
2. **In-line Prevention:** Digital authentication of photos/videos at the time of capture to “tamper-proof” the media at the point of capture. This could simply be as part of a secure app that prevents the insured from uploading their own photos, or even better, utilizing a blockchain or immutable ledger that protects against both inside and outside changes to the media, by utilizing a global consensus model.

Diving into further detail on the two approaches, detection does have a few disadvantages. These include the amount of time and processing required to analyze photos or videos. Extensive analysis using AI is challenging to run in-line as photos are gathered from claims. Additionally, this analysis may be a never-ending cat and



stock.adobe.com: only4denn; Fotomay

> “DEEPFAKES OR SYNTHETIC MEDIA CAN EFFECTIVELY BE USED TO FILE FRAUDULENT CLAIMS, CREATE FRAUDULENT INSPECTION REPORTS, AND EVEN ESTABLISH THE EXISTENCE AND CONDITION OF ASSETS THAT DO NOT EXIST.

mouse game, similar to virus scan, given constant improvements in deepfake technology. The detection tools designed to flag manipulation will always be chasing, evolving, and improving editing tools that do the manipulation.

On the other hand, detection is sometimes the only defense when the media is not captured by a trusted application or trusted person. For instance, if an insured sends claim photos via email, an insurer has only two options: Request the photos are retaken from a trusted application or accept the photos and perform an analysis to ensure the photos are authentic. To answer that question, unless an insured has a record of insurance fraud, a bad claims experience is unlikely a reasonable tradeoff for better fraud reduction.

That brings us to prevention technologies, which unlike detection offer a more reliable and future-proof solution to the deepfake problem. By locking down the media at the point of capture, so that any changes become tamper-evident, we are much more confident that the content we are viewing is original and unchanged. Think of it as digitally watermarking that doesn’t necessarily scribble on the photos. The one catch is prevention only applies at the point of creation or capture, which means it cannot always replace detection as the best defense when capture software is not available or not used.

Pragmatically, this may suggest a hybrid arrangement, starting with a secure app that captures claims photos, authenticating them at the point of capture. Now assuming all insureds use the app, then the threat of deepfakes is removed. Outside of this ideal world, we know app adoption is not 100% and ultimately, some claims will seep through via other less secure processes requiring some form of detection.

The verdict? Some carriers may try to push higher adoption of their apps with in-line security, others may choose to take a hybrid approach of prevention and detection, while others may just assume the risk of additional fraud, relying on the discouragement already in place through criminal prosecution of fraud and in the hope that standards soon emerge to prevent deepfakes and synthetic media from impacting their claims.

Insurance is vital and complex, safeguarding all aspects of our lives, but it’s also increasingly vulnerable to new methods of fraud and deception with the emergence of deepfakes. With the growing adoption of self-service and ways in which digital media can be easily compromised, it’s critical to begin to challenge the status quo within fraud prevention while leveraging future standards once they become available. Steps you take to protect insurance claims today will continue to pay off in years to come.

Nicos Vekiarides (nicos@attestiv.com) is the chief executive officer & co-founder of Attestiv. He has spent the past 20+ years in enterprise IT and cloud, as a CEO & entrepreneur, bringing innovative technologies to market.

Reprinted with permission from the September/October 2021 edition of Claims © 2021 ALM Media Properties, LLC. All rights reserved. Further duplication without permission is prohibited. For information, contact 877-256-2472 or reprints@alm.com. CLM-10202021-524346