IDENTITYWEEK

GLOBAL • TRUSTED • VISIONARY

THE NEXT CHALLENGE: THE GOLD STANDARD FOR DEFEATING AI DEEPFAKES

EVIE KIM SING
EDITOR
IDENTITYWEEK.NET

FEBRUARY 2024

Introduction

The principal threats to biometric technology, such as face-matching and presentation attacks are now widely considered to be legacy problems - i.e. Are they the right person or a real person/artefact?

That was yesterday.

Today, the big threat comes from synthetically generated imagery (AI) digitally injected into data streams to create false media which is indistinguishable from a real identity. These false images may be used to facilitate passport and further document fraud, illegal migration, human trafficking, non-consensual intimate media, and more illicit activities.

The influx of fake imagery of this sophistication - called deepfakes or morphed images - represents the current challenge that society must deal with.

This report compiles research and presents the results of an expert survey on deepfake and morphing trends that was conducted in January 2024. Selected experts from the ecosystem were invited to respond.

The problem of AI generated deepfake fraud

The EU's impending AI Act has undergone extensive scrutiny and deliberation within EU institutions since 2018, due to the argument for regulating AI versus enabling innovation. In this way, the development of artificial Intelligence is a double-edged sword. The Act is set to become the EU's first sweeping AI law in 2024, adopting a tiered-approach to restrictions based on the risk assessment of technologies. While society has explored many innovative use cases using AI, fraudsters utilise AI to counterfeit physical and digital documents. It is easy to create a deepfake and harder to spot the signs.

The fraud landscape is evolving with more sophisticated methods from morphing to advanced morphed imagery, called deepfakes.

iProov, a leading biometric authentication company, defines deepfakes as "a video, visual, or audio recording that has been distorted, manipulated, or synthetically created using deep learning techniques to present an individual saying or doing something that they did not say or do."

"Deepfakes are created using artificial neural networks".

A recent report by Sumsub shows a 10x increase in deepfake attacks worldwide from 2022 to 2023, making this the most trending and prolific type of fraud.

The MEA region has seen an increase in deepfakes of 450%.

The prevalent use of AI enables anyone to be a victim of realistic photo or video imitations. Common attacks by AI impersonators include targeting onboarding processes for digital banking or the global travel industry. Deepfakes can usually bypass any identity checks to create a false account for money-laundering. The advanced forgery can affect both digital and physical documents.

Prosperous economies in particular in the European Union (and the UK) will attract fraudsters and bolder attacks, the research suggests.

This is why the half-and-half debate for prohibitions proposed in the AI Act has been turned around in favour of labelling technologies according to their levels of risk.

NIST face recognition vendor testing & countering the threat

AI-powered biometrics is a tool helping to detect deepfakes as well as video authenticator software, neural networks, autoencoders and forensic techniques.

iProov's flagship technology, Genuine Presence Assurance, combines legacy actions for organisations such as identifying if the person authenticating is the right person and real, and insights into real-time only attacks.

Pavel Goldman Kalaydin, Head of AI/ML at Sumsub, says being vigilant of how deepfakes may be presented is very important, whether in an advertising campaign, electoral speech or identity document. This is an important step as the creation of deepfakes is premeditated, as it requires time and expertise to produce a convincing deepfake image or video.

TECH5 is a leading developer of biometric algorithms which consistently rate in the top tier of NIST industry benchmarks for facial, iris and fingerprint matching. NIST performs independent testing of prototype face morphing attack detection (MAD) technologies designed to obtain a common performance assessment. Open to participation worldwide, the FRVT MORPH test has provided recognised rankings for solutions since June 2018.

"Face morphing is an image manipulation technique where two or more subjects' faces are morphed or blended together to form a single face in a photograph"

Face Recognition Vendor Test (FRVT) Part 4: MORPH - Performance of Automated Face Morph Detection

NIST's Vendor Test assesses the performance of automated face morph detection technologies to combat some of the biggest challenges, including printing or scanning photos digitally into a photo document i.e. a passport.

Some morphing techniques use a single image provided to the algorithm or two-image differential morphing is where two face images - a morph and a real face image - are merged to generate a new identity.

Vendors are using a "blend of certified management and security systems, like ISO 27001, certified products e.g. eIDAS and NIST 800-63B and AI-powered biometrics, detection algorithms", says **Haraldur Bjarnason, CEO, Auðkenni.**

"I don't think there are vendors that specifically develop technology for (deepfake) defects out there.. There will always be some sort of new arms race, so people try to come up with better defects".

"Platforms invest heavily in technology combating deepfakes and devices for verification and signals. Technology and regulation should be complimentary."

Pavel Goldman Kalaydin, Head of AI/ML, Sumsub

Effect: The creation and distribution of counterfeit documents

The creation and distribution of counterfeited documents must be better controlled to maintain customer relationships, commercial rates, and security. The issue that fraudsters are making legitimate applications to the U.S. Citizenship and Immigration Service (USCIS) and other entities using stolen identities and fraudulent photographs is a serious concern. Criminal networks seek the issuance of counterfeit documents to commit immigration fraud and engage in human trafficking across borders. However, according to Sumsub, organisations themselves - who just want to identify and verify the user - are not aware of the large scale of the problem.

AI standards & regulation

The EU AI Act will be implemented in 2024 with high expectations for it to be a comprehensive law imposing accountability, transparency, and risk management for all artificial intelligence technologies, including deepfakes. With the array of technologies becoming more accessible to all, deepfakes are an attractive method for fraudulent activity and easy to create.

The European Parliament specifies that published content must come accompanied by a watermark or labeling to prove authenticity.

The regulation environment around AI is also supported by ISO standards, but in relation to deepfakes, standards bodies like ETSI have published delivery reports like, ETSI GR SAI 011 Securing Artificial Intelligence (SAI) and Automated Manipulation of Multimedia Identity Representations. The international standard (BS ISO/IEC 4200) by BSI, the UK's National Standards Body, sets out how to implement and maintain an AI management system. More development is expected around standards and regulation on this topic in the forthcoming years.

Securing Artificial Intelligence (SAI) addresses 4 ways of AI standardisation

- Securing AI from attack to generate deepfakes
- Mitigating the 'problem' of AI (misuse)
- Using AI 'solutions' for detection e.g. AI-powered biometrics
- Safety and security aspects of AI

The 2023 International Standards Organisation published global guidance on how companies can introduce an AI system, which will "get them closer to being compliant with the EU AI Act", according to Sumsub.

Across global regions, China's state power is not introducing the toughest restrictions as some may expect, but will gradually adopt a concrete "Artificial Intelligence Law" like the European Union which standardises healthy AI technologies. In the United States, the 'No AI Fraud Act' will eliminate the misuse of deepfakes.

Digital twins, deepfakes and defamation of public figures

The commoditisation of AI has led to many in the acting profession wanting to leave a lasting legacy in Hollywood by creating their own digital twin. However, notable public figures like Tom Hanks, Scarlett Johansson, Taylor Swift, and Mr. Beast have also fallen victim to deepfake identity cloning. The result is fraudsters are modifying AI deepfake images or videos of a celebrity to make money from advertising campaigns without their consent.

Companies like HAND (Human & Digital) work with the C2PA.org to provide provenance attestations via unique identifiers as part of a composable, multi-factor authentication model. HAND creates a global ID registry for talented digital twins linked to their legal identities.

Conclusion

Overall, the focus of international regulation is to encourage companies to develop, implement and use standardised, healthy AI models. Restrictions led by the European Union's comprehensive AI Act will only be deemed necessary for the top tier of 'high risk' technologies. The tier system favours a less restrictive approach for the majority of AI creations contributing to a better society. AI has been impactful in so many positive ways, in particular for the development of age predictive portraits of missing people.

The impact of deepfakes is currently felt by organisations rather than the customer or public, who must learn to spot the signs of deepfakes reemerging in new situations. There is heavy investment in biometric authentication and verification methods by vendors like iProov and TECH5, which the report highlights.

With thanks to our contributors

Will Kreth, CEO, **HAND (Human & Digital)**
Haraldur Bjarnason, CEO, **Auðkenni**
Pavel Goldman Kalaydin, Head of AI/ML, **Sumsub**
Natalia Fritzen, AI Regulation expert, **Sumsub**

Appendix

GR SAI 011 - V1.1.1 - Securing Artificial Intelligence (SAI); Automated Manipulation of Multimedia Identity Representations (etsi.org)

Face Recognition Vendor Test (FRVT) Part 4: MORPH - Performance of Automated Face Morph Detection (nist.gov)

Artificial Intelligence Act: deal on comprehensive rules for trustworthy AI | News | European Parliament (europa.eu)

sumsub.com/files/sumsub_identity_fraud_report_2023.pdf

Four things to know about China's new AI rules in 2024 | MIT Technology Review

ETSI - Best Security Standards | ETSI Security Standards

Key developments proposed for the EU AI Act as it moves to latter stages | Deloitte UK

iProov | Authenticate & Onboard Online Users With Biometrics

Home - TECH5: Innovative Identity Management Technologies

Identity Verification Service - G2 Leader 2023 | Sumsub

identityweek.net