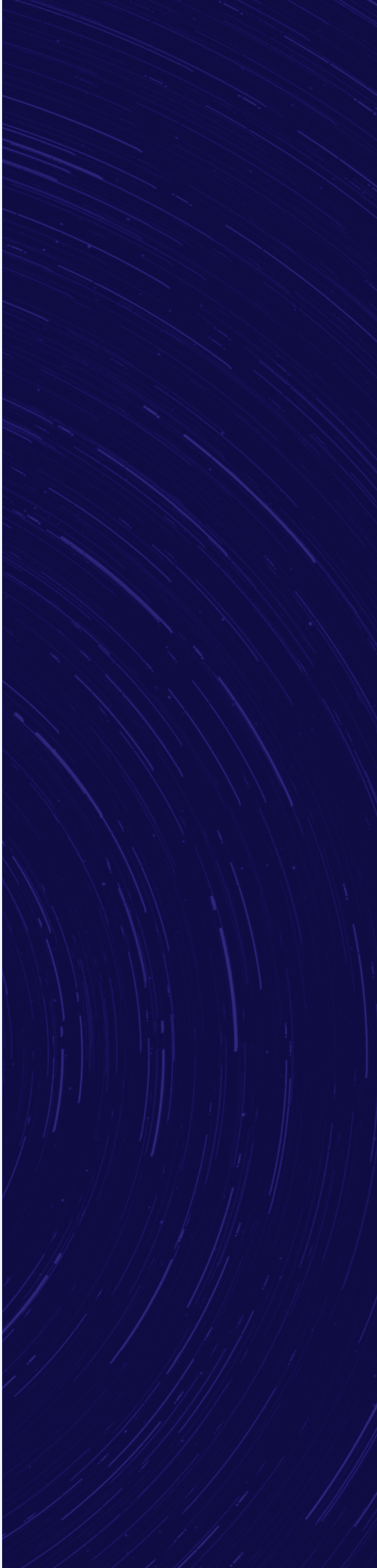
The State Of Deepfakes 2024

sensity



Quick note

Since we started Sensity AI in late 2018, our focus has always been on tracking the threat of malicious use of Generative AI techniques in videos, images, and audio in online environments.

Our continuous research aimed not only to understand the evolution of threats across different industries but also to gather intelligence about which countries were most likely to be targeted by AI-powered cyber threats and which were more likely to be vectors and outlets for these menaces. Consequently, we were able to focus on what mattered most at a specific time and location by providing government agencies and companies with cutting-edge technologies to detect and respond. After the COVID-19 outbreak in 2020, the world was forced to shift from physical to fully digital interactions.

This sudden acceleration has undoubtedly benefited the overall process of digitizing economies and societies, but it has also sped up the creation of new cyber threat techniques, frauds, and new ways of influencing large swathes of the population by leveraging Generative AI technologies that have boomed in the past four years.

Now, anyone without technical skills can generate images, videos, and text by simply describing the desired outcome, or clone facial and audio biometrics and voice of any human being on Earth from a single photo and a vocal message. Cybercriminals, hacktivists, adversarial countries, fraudsters, fake news outlets, and cyber soldiers have quickly incorporated AI technologies into their attack and deception frameworks, faster than anyone in the public and private sectors expected. In this report, we want to highlight the current state of deepfakes in the real world beyond alarmism, buzzwords and hype-driven topics spreaded online.

Our goal is to show where bad actors are operating, in which geographies they act, what techniques and frameworks they use, who their targets are and which technologies they are using. We also focused on ongoing events like the USA 2024 Elections, the Russia-Ukraine war, and the Israel-Hamas conflict.

We will leverage a vast amount of intelligence collected from our partners over the past years, carefully anonymized, to assess the situation and understand the impact this new generation of cyber threats is having on ongoing events and to enable governments, NGOs, companies, and citizens to grasp the extent of the threats and prepare for future challenges.

Francesco Cavalli

Co-Founder
COO & Chief of Threat Intelligence

Is Generative AI fully commodified?

By 2024, Generative AI has fully integrated into the fabric of multimedia creation, transforming the landscape of video, image, and audio production.

The once complex and resource-intensive processes have been democratized, allowing creators at all levels to harness sophisticated AI tools with unprecedented ease. From automated video editing and realistic image generation to advanced audio synthesis, the technology has become fully commodified, enabling rapid, high-quality content production. This commodification marks a significant shift, where the barriers to entry are significantly lowered, fostering a new era of creativity and innovation in digital media.

As mentioned in the previous pages, the commodification of Generative AI also brings significant cybersecurity challenges. Identity theft, scams, face and voice biometrics cloning, but also misinformation campaigns and reputation attacks are growing worldwide.

2.298

TOOLS FOR AI FACE SWAP, LIP SYNC, FACE REENACTMENT, AI-AVATARS

10.206

TOOLS FOR AI IMAGE GENERATION

1.018

TOOLS FOR AI VOICE GENERATION, VOICE CLONING

47

TOOLS FOR DEEPFAKE KYC INJECTION

Some numbers accessible by anyone on the internet. The math includes: public repositories, open source projects, free tools and paid tools priced less than $50/ month.

Introduction

In this work we reflect on the evolving landscape of digital media manipulation in 2023 and first half of 2024.

As public awareness and technological defenses against deepfakes have improved, adversaries have adjusted their strategies accordingly. They are increasingly turning to sophisticated deepfake technologies that allow for the creation of more convincing and difficult-to-detect synthetic media. This shift aims to exploit trust and manipulate reality at a pace that challenges current detection capabilities, underscoring the critical need for continuous advancement in both detection technologies and digital literacy education.

These techniques are evident in the alarming rise of deepfake-related incidents, a highly disruptive and potentially lucrative endeavor for malicious actors. Unsurprisingly, the misuse of deepfake technology persisted as one of the most pervasive threats in the digital landscape of 2023. Adversaries have refined their methods to maximize deception, realism, and psychological impact, leveraging advanced techniques to create and disseminate convincingly altered imagery, videos and audio recordings that challenge the boundaries of trust and authenticity across several industries.

Several different entities were similarly active in the realm of deepfake technology throughout 2023 and first half 2024. These efforts often involved creating synthetic videos, imagery or audio aimed at discrediting political figures or manipulating geopolitical narratives, but also stealing identities and spread extremely well built scams impersonating public figures worldwide.

Bad actors classification in the deepfake threat landscape

CYBER CRIMINALS

They use deepfakes to commit fraud and extortion, manipulating videos and audio to create false evidence or deceive individuals.

CYBER SOLDIERS

Employed by nations or organizations, these actors use deepfakes to conduct cyber warfare, phishing attacks and spreading disinformation or to destabilize internal or geopolitical situations.

ADVERSARY AGENCIES

Intelligence or defense agencies deploy deepfakes as part of their cyber paramilitary operations, using sophisticated technology to undermine the political, social, and economic stability of rival countries or just influence public opinion.

FAKE NEWS OUTLETS

They specialize in the creation and dissemination of deepfake content to mislead the public, sway opinions, and create divisive narratives in society.

FRAUDSTERS

These individuals exploit AI technology to steal face and voice biometrics in order to bypass KYC facial recognition and liveness check, often targeting fintech and digital banking applications.

HACKTIVISTS

They use deepfakes to advance political or social agendas, creating manipulated media to support their causes or attack their adversaries, blurring the lines between activism and cybercrime.

Threat Landscape

Influence Campaign

Deepfake influence campaigns represent one of the most impactful types of destabilization instruments for influencing public opinion ever designed, as the commodification of AI is growing, these campaigns are becoming easier to create and more effective.This manipulated content can be targeted and disseminated through social media and other digital platforms, aiming to amplify societal divisions and undermine the stability of political, economic, or social structures within a country. In the past 5 years we have seen an exponential employment of deepfakes as part of hybrid war operations especially in East Europe, Asia and United States.

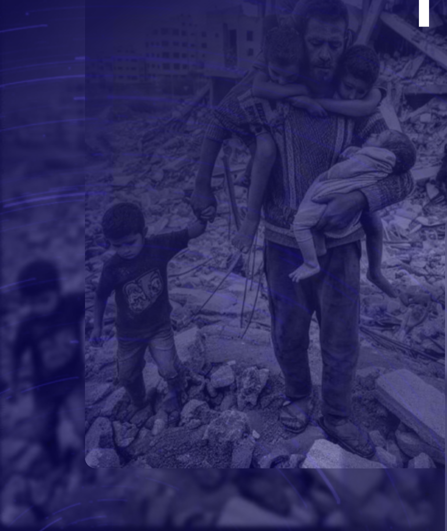
Scams

Deepfake scams employ convincingly altered videos and audios to mimic credible entities or individuals, tricking people into revealing personal data or making fraudulent payments. Spread via social media and digital ad networks, these scams leverage the extensive reach and precise targeting of these platforms to engage and exploit a large number of victims for financial profit. Outside the political or geopolitical space, this is the most dangerous threat to social media users given the realty level that deepfake technologies are guaranteeing to scammers such as hyper realist lip sync and voice cloning, scam size can go from 200$ up to tens of thousands of dollars depending on the target victim capabilities. These scams are easy to create and spreaded generating huge returns for cyber criminals and cyber soldier groups. Ad networks turned out to be an important and scalable help for maximizing the effect of these scams.

KYC Frauds

KYC frauds involving deepfakes use sophisticated video and image manipulations to defeat facial recognition and liveness detection systems used in digital onboarding processes. By presenting forged biometric data that appears genuine, fraudsters can bypass security measures, enabling unauthorized access to financial services and sensitive personal accounts. Based on the evidence gathered in the past 3 years, fraudsters groups and individuals in LATAM are mastering this technique very well and effectively exploiting the lack of security countermeasures in the whole region. South East Asia and Africa are being heavily impacted by these particular attacks

Deepfake Influence Campaigns



INFLUENCE CAMPAIGN

Deepfakes have emerged as a potent tool in influence campaigns, altering the landscape of information dissemination and public perception.

These sophisticated digital creations are being utilized with increasing frequency to engineer consent or dissent, reshaping the dynamics of both political landscapes and international relations.

The ability to seamlessly manipulate audio and video content allows entities (ranging from political operatives to state-sponsored groups) to fabricate scenarios or statements with the goal of deceiving the public, influencing elections, or even destabilizing geopolitical adversaries.

This convergence of advanced technology and strategic misinformation represents a significant challenge to the integrity of communications.

HACKTIVISTS

ADVERSARY AGENCIES

FAKE NEWS OUTLETS

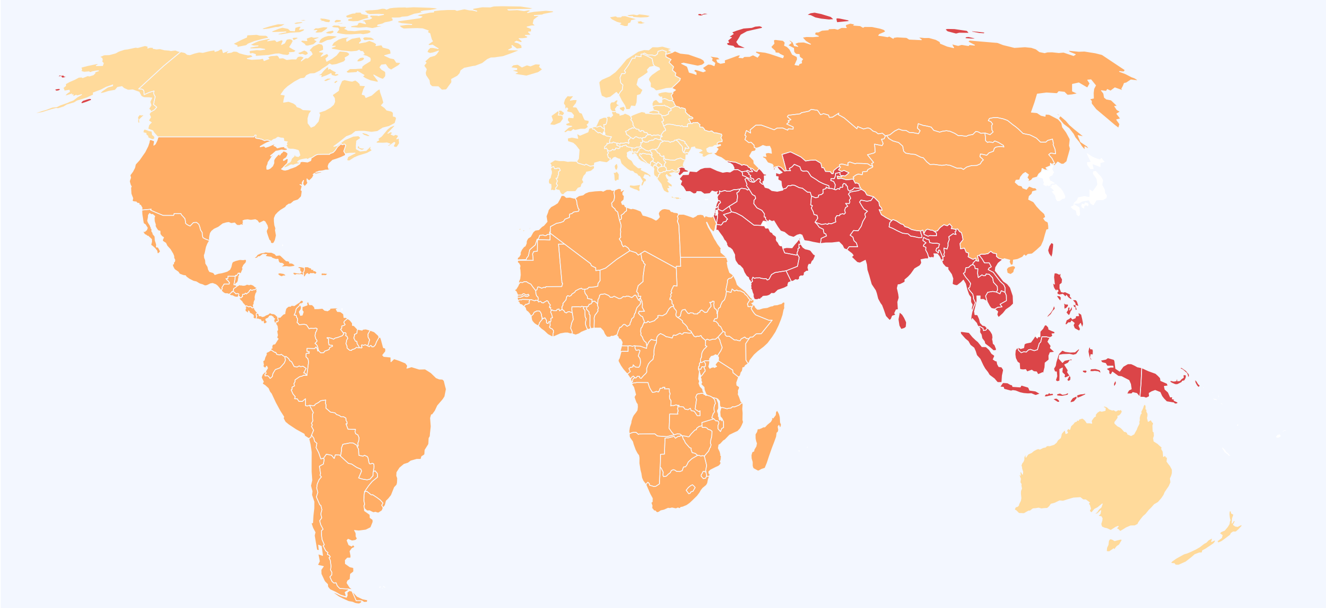
CYBER SOLDIERS

What, Why. Where, Who

	WAR SCENARIOS	POLITICAL CONTEXT
WHAT	Deepfakes may go beyond targeting individual reputations to encompass broader deceptions, such as faking news reports of military actions that never occurred, creating fake endorsements from influential figures, or simulating atrocities to garner international support or condemnation.	In these settings, deepfakes often depict politicians or public figures in misleading or false situations—making inappropriate comments, taking controversial stances, or appearing in fabricated locations.
WHY	In these contexts, deepfakes aim to undermine the enemy's stability, create confusion among its citizens and allies, and spread disinformation to weaken opposition without engaging in direct military conflict.	Deepfakes are used to manipulate public opinion, discredit political opponents, and influence election outcomes. The goal is to create false narratives or exaggerate truths to sway voter behavior and preferences.
WHERE	They are disseminated across similar digital platforms but are often tailored to disrupt or mislead specific populations or military groups, sometimes even targeting international audiences to influence global perspectives on a conflict.	These deepfakes are primarily spread through social media platforms, online ads, and digital news outlets to target specific voter groups and demographics, maximizing emotional impact and belief propagation.
WHO	State actors, intelligence agencies, and sometimes third-party groups aligned with governmental objectives engage in the production and strategic deployment of deepfakes, using them as an element of broader hybrid warfare tactics.	Political movements and politicized citizens are the primary creators and distributors, utilizing deepfakes as a modern tool for discrediting political opponents or amplifying the prestige of one's own candidate.

Worldwide distribution

The map below shows the distribution of generative AI techniques used for influence campaigns by region.



🟥 Red

Regions where deepfake influence campaigns are having **extremely high impact** on public opinion

🟧 Orange

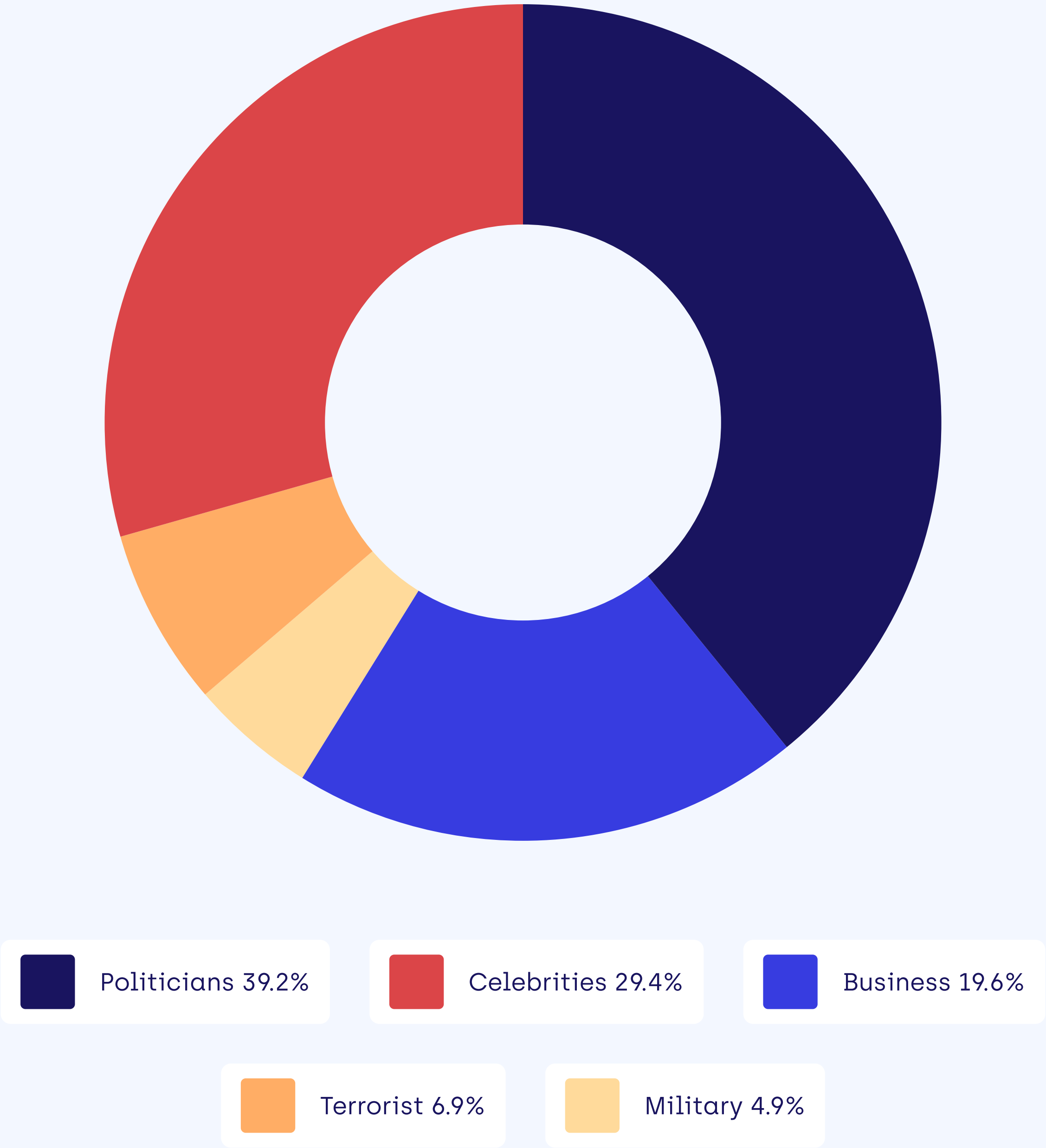
Regions where deepfake influence campaigns are having a **significant impact** on public opinion

🟨 Yellow

Regions where deepfake influence campaigns are having a **limited impact** on public opinion

Targets

The chart shows deepfake influence campaigns targets



Anatomy of deepfake influence campaigns

Intelligence gathering

The initial phase involves a thorough analysis of the target state's political environment, cultural aspects, media consumption patterns, influential figures, and prevailing public opinions. This step integrates multiple intelligence collection methods, including human intelligence (HUMINT), signals intelligence (SIGINT), open-source intelligence (OSINT), and social media monitoring tools. These resources are pivotal in acquiring a nuanced understanding of the target landscape, which informs subsequent strategies.

Target Audience Analysis

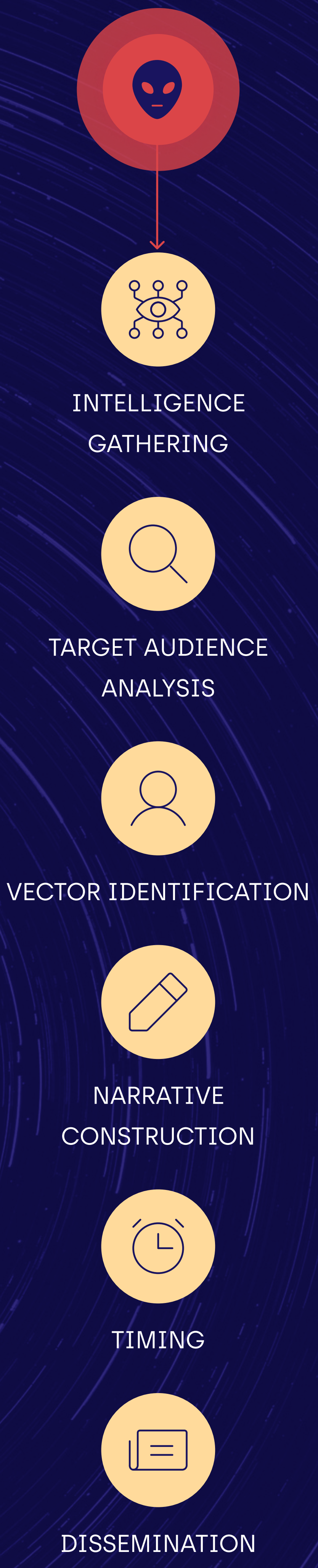
These groups are selected based on their susceptibility to influence and the likelihood of their receptivity to tailored messaging. A detailed examination of their beliefs, preferences, and online activities is crucial to customize content that resonates deeply and persuasively with them.

Vector identification

Often recognizable public figures relevant to the adversary's context. These figures can be leveraged to enhance the credibility and reach of the campaign, acting as conduits for the dissemination of crafted messages.

Narrative Construction

The narratives are designed to align with the campaign's objectives, which could range from undermining political figures to exacerbating societal divisions, eroding trust in governmental institutions or amplifying internal propaganda. It is vital that these narratives are coherent, emotionally engaging, and crafted to elicit specific responses from the audience particularly susceptible to AI-generated media.



Timing

The release is timed to maximize impact, ideally aligning with significant events, political debates, or periods of increased societal tension within the target state. Sustaining a consistent output of content is also necessary to maintain engagement and reinforce the campaign's messages over time.

Dissemination

This involves deploying a sophisticated network of bots, trolls, and fake accounts across various digital platforms including social media, online forums, and digital news outlets. Preferred methods are targeted advertising, hashtag hijacking, and coordinated sharing efforts. These tactics are designed to maximize the content's reach and visibility, ensuring that it penetrates the intended audience segments deeply and broadly, thus amplifying the impact of the campaign's messaging.

Timing - Closer look

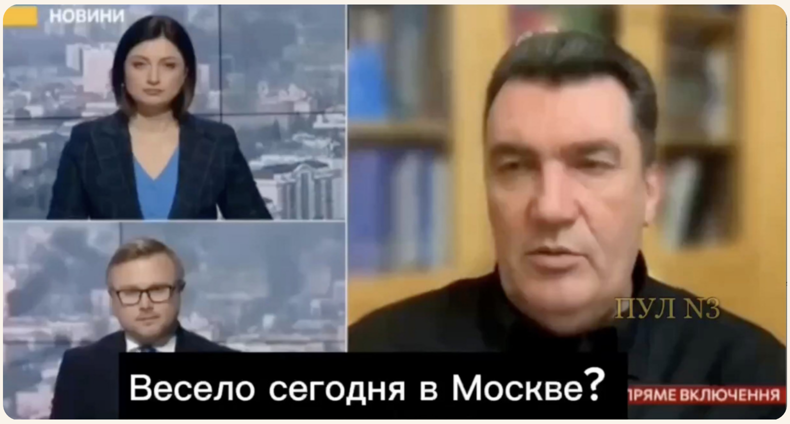
After the recent terrorist attack in Moscow the russian propaganda released a deepfake about the Secretary of the National Security and Defence Council of Ukraine Oleksii Danilov admitting the ukraine responsibility behind the attacks in an alleged official TV interview with real journalists. The timing here played a crucial role, it was released and disseminated on russian TV and socials in less than 24h supporting the first rumors from the Kremlin that Ukraine has somehow partecipated to the attacks by offering refugee to the terrorists. In a very controlled media environment like the russian one, the deepfake has certainly influenced millions of people that do not have access to international debunking services.

WATCHLIST

Influence Campaigns

Deepfake Influence Campaign Samples

See ↗



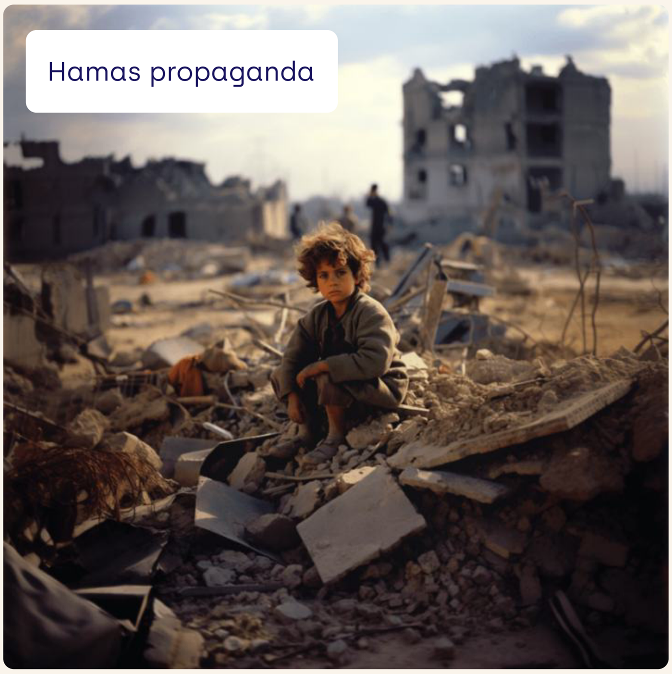
SPOTLIGHT

Israel - Hamas conflict

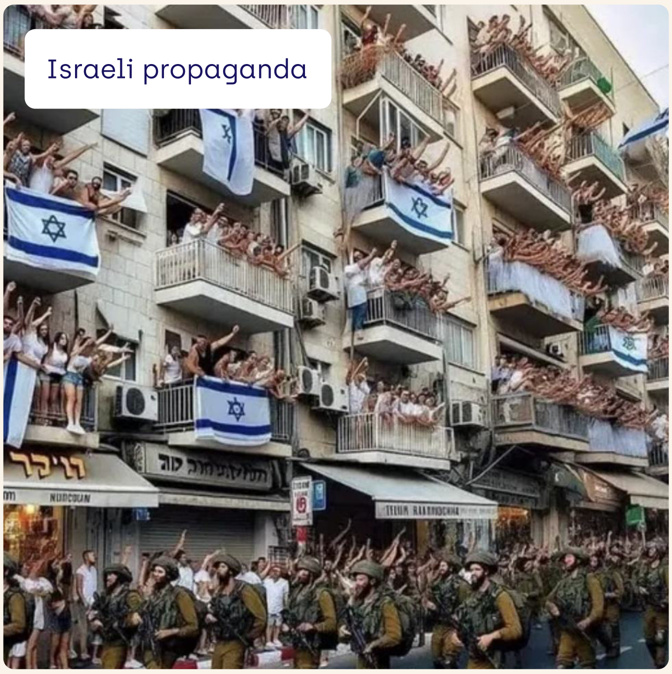
Between the attack on October 7, 2023, and the first weeks of the IDF (Israeli Defense Forces) offensive, we have witnessed a massive propaganda conflict on social media between the two opponents to gain public opinion support worldwide.

This media confrontation saw extensive use of all kinds of AI-powered tools to create narratives favorable to their side in order to influence as many people as possible in the shortest time possible.

On one hand, Hamas propaganda made extensive use of AI-generated images depicting Gaza residents (especially children) next to the rubble of homes with dead and injured people. On the other side, Israeli propaganda focused on internal efforts to unite the country using generated images depicting huge crowds with Israeli flags and military personnel in parades, as well as deepfake videos, combining lip sync and voice cloning, portraying global celebrities expressing their support for Israel.



Hamas propaganda



Israeli propaganda

SPOTLIGHT

The US 2024 elections

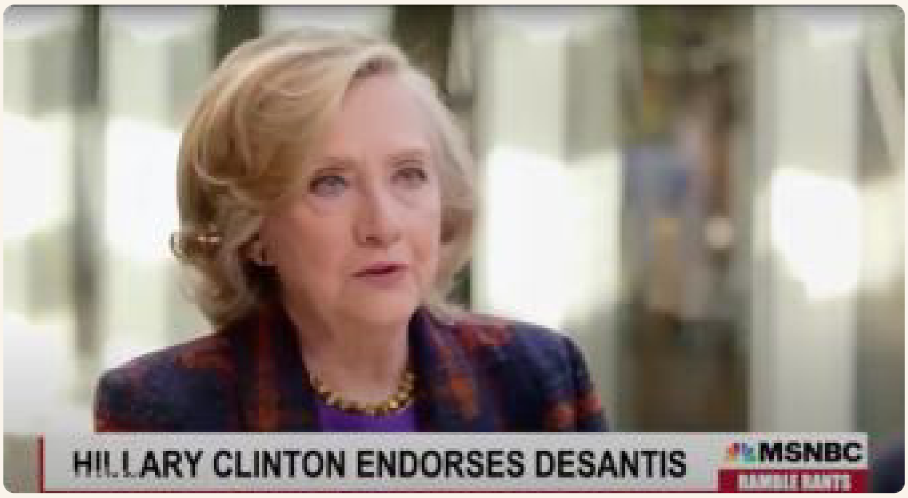
The 2024 U.S. presidential election is crucial in shaping the global geopolitical landscape. With the world facing numerous challenges, including the Russia-Ukraine conflict, rising tensions with China, and instability in the Middle East, the election's outcome will significantly impact international relations and global strategies.

The U.S. election is closely watched by major global players like Russia, China, Iran, and North Korea, each with vested interests in the future direction of American foreign policy.

To prevent the information poisoning witnessed during the 2016 U.S. election, numerous initiatives have been implemented ahead of the 2024 election.

The Cybersecurity and Infrastructure Security Agency (CISA) has launched the #Protect2024 campaign, which includes measures such as multi-factor authentication, cyber hygiene vulnerability scanning, and physical security assessments for election offices. Additionally, CISA's #TrustedInfo2024 initiative aims to promote election officials as reliable sources of information, countering misinformation and foreign influence operations (CISA).

Although the election campaign is still in an early phase we have found initial evidence of deepfake weaponization during the primary election in particular against the main Donald Trump opponents.



SPOTLIGHT

The US 2024 elections

The Ron DeSantis Case

Over the past seven months, 48 deepfake videos featuring Ron DeSantis were detected, with the peak of this campaign coinciding with the onset of the Republican Party primary elections. These deepfakes, which included fabricated videos of DeSantis making various announcements and statements, spread widely on social media platforms such as Twitter and TikTok.

One prominent example was a video that falsely showed DeSantis announcing his withdrawal from the presidential race, which gained significant traction and was viewed thousands of times before his actual retirement from the race.

Watchlist

See ↗

5.8M

VIEWS

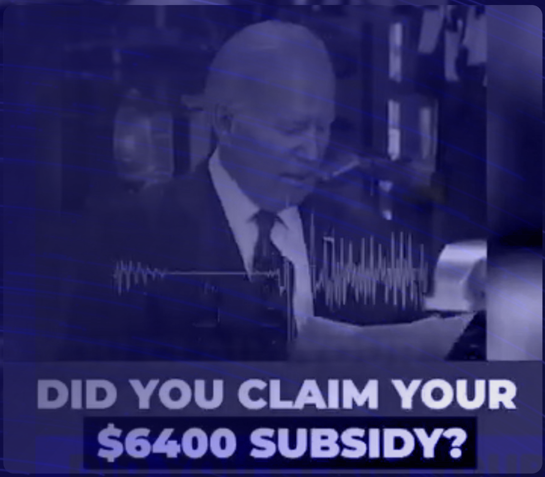
32K

INTERACTIONS

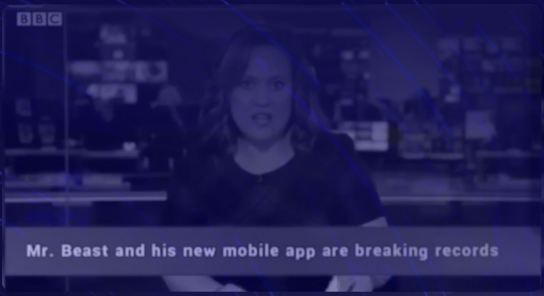
12

SOCIAL MEDIA PLATFORM/WEBSITES

Influence campaign results



Deepfake Scams



What, Why. Where, Who

SOCIAL MEDIA SCAMS

WHAT

These deepfakes are designed to deceive users into believing that high-profile individuals endorse these potentially lucrative opportunities, which are, in reality, fraudulent schemes.

WHY

The primary motivation behind these scams is financial gain. Scammers use deepfake technology to lend credibility to their schemes, exploiting the trust and admiration that potential victims have for the figures being impersonated.

WHERE

These scams are predominantly found on social media platforms such as Facebook, Twitter, Instagram, and YouTube, where it's easy to reach a large audience quickly and where users are more likely to share content virally. Social media also allows for targeted advertising, enabling scammers to reach specific groups more likely to be interested in trading, gambling, or crypto.

WHO

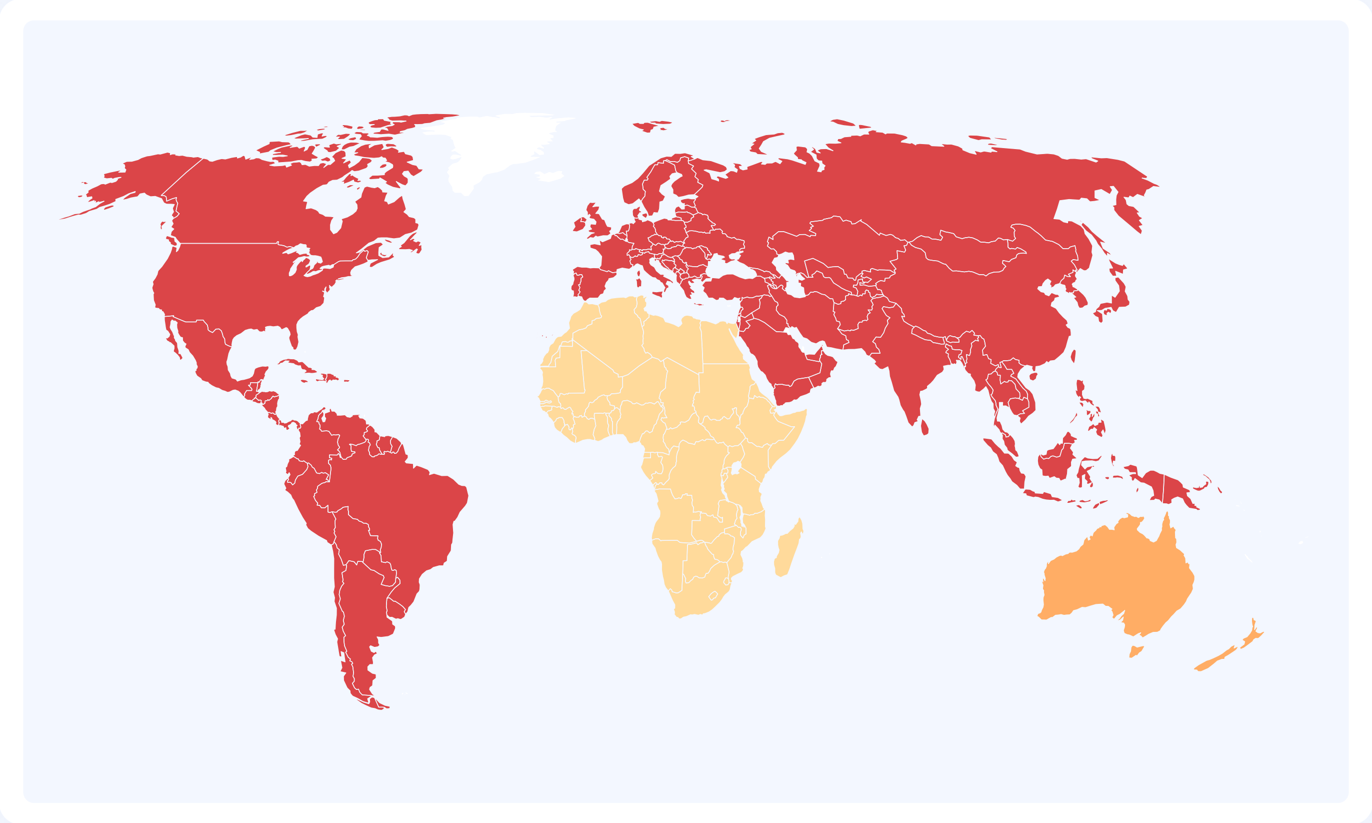
The perpetrators behind these scams are typically organized cybercriminal groups with enough skills to create convincing deepfakes and understanding of social media algorithms to effectively disseminate their content. They often operate in jurisdictions, making it challenging to track and prosecute them.

CYBER CRIMINALS

FRAUDSTERS

Worldwide distribution

The map below shows the percentage of generative AI techniques used for Social Media and Ad Networks scams by region.



🟥 Red

Extremely high number of deepfake scams circulating on Social Media and Ad Networks

🟧 Orange

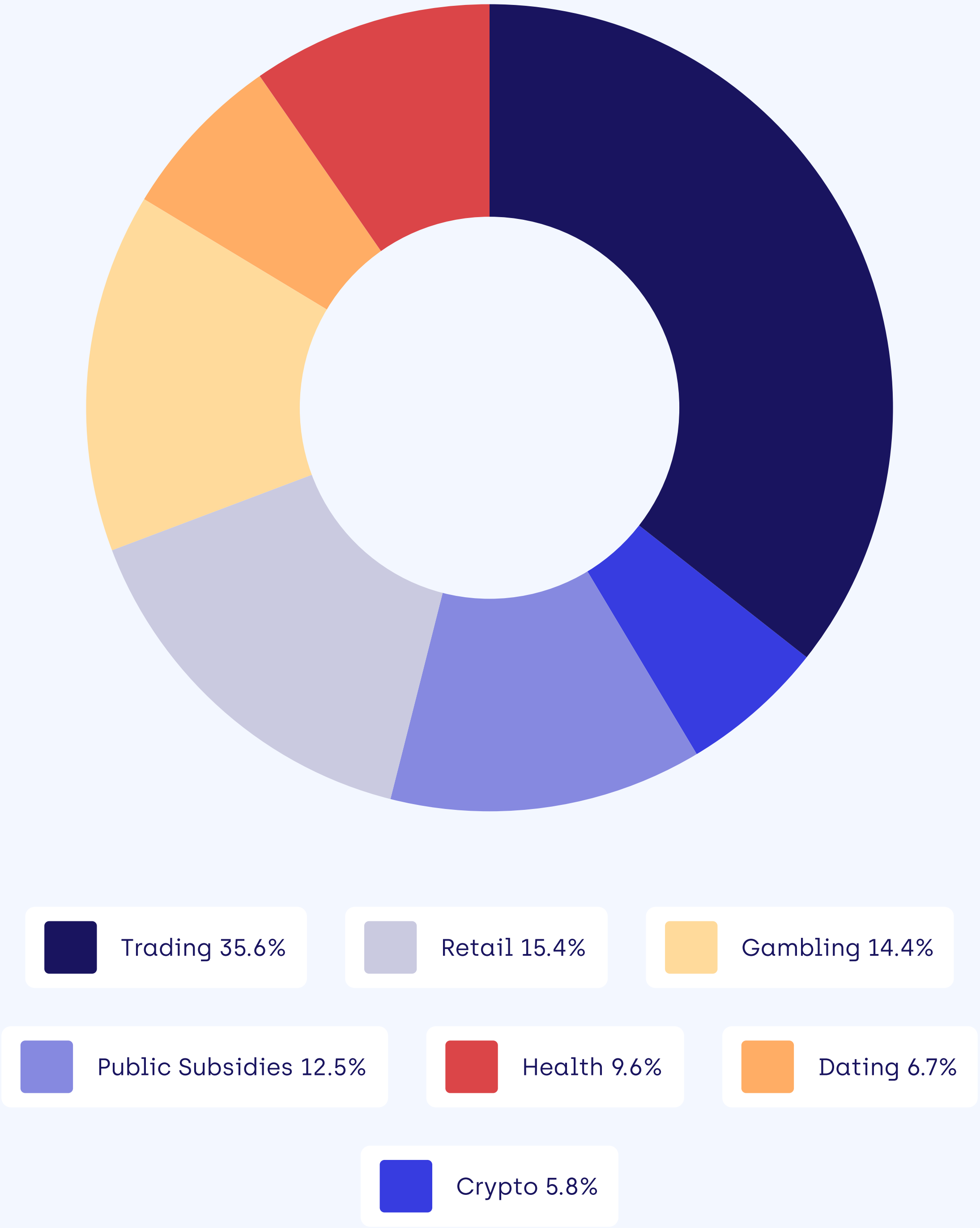
High number of deepfake scams circulating on Social Media and Ad Networks

🟨 Yellow

Limited number of deepfake scams circulating on Social Media and Ad Networks

Most targeted industries

The chart shows deepfake scams most targeted industries



Anatomy of a deepfake scam

Goal Setting

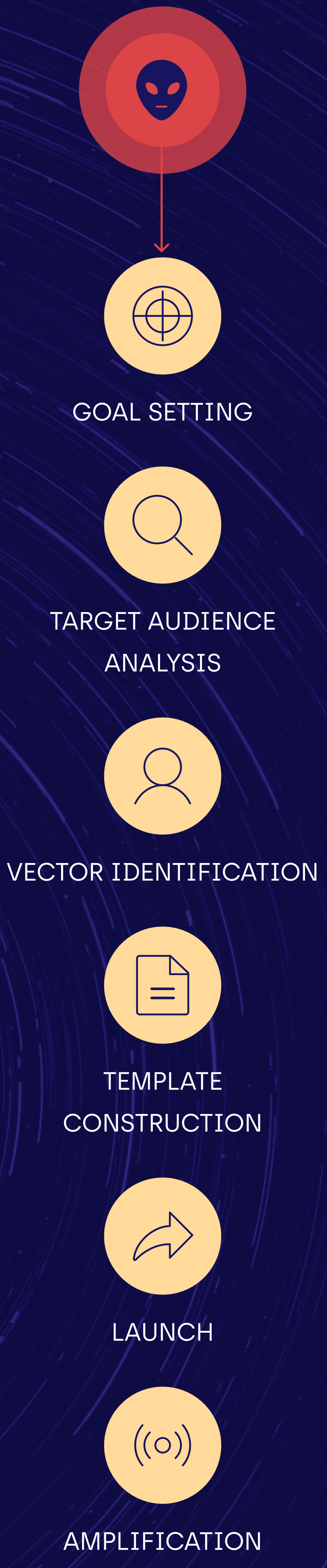
The goal is typically centered on financial gain through deceptive means. Scammers aim to create realistic deepfake videos or audio clips that portray high-profile individuals endorsing fraudulent schemes or products. The overarching objective is to exploit the trust and admiration viewers have for these figures, convincing them to invest money or share sensitive information.

Target Audience Analysis

Scammers conduct thorough analyses to identify the most susceptible segments of social media users. They may target users known for following specific celebrities, influencers, or financial gurus, or those who have shown interest in areas like investments, cryptocurrency, or high-return schemes. By understanding the demographics, interests, and online behaviors of these groups, scammers can craft more persuasive and targeted scams.

Vector identification

This phase involves selecting the most effective platforms and methods for deploying the scams. Given that these scams thrive on broad exposure and viral sharing, popular social media platforms like Facebook, Twitter, Instagram, and YouTube are common choices. These platforms not only allow scammers to reach millions quickly but also offer sophisticated targeting options through ads and algorithmic content distribution.



Template Construction

At this stage, scammers develop the templates for their scams, which include creating the deepfake videos or audios, crafting accompanying texts or posts, and designing any supporting fake websites or contact points. The templates are made to look as authentic as possible, often mimicking the style and tone of the impersonated celebrities or the media formats commonly used on the chosen platforms.

Launch

In this phase the scam is released to the public, typically through social media posts, sponsored ads, or direct messaging. The initial release is strategically planned to maximize reach and engagement, possibly timing it to coincide with relevant events (like a new product launch by the impersonated celebrity) to piggyback on existing media buzz.

Amplification

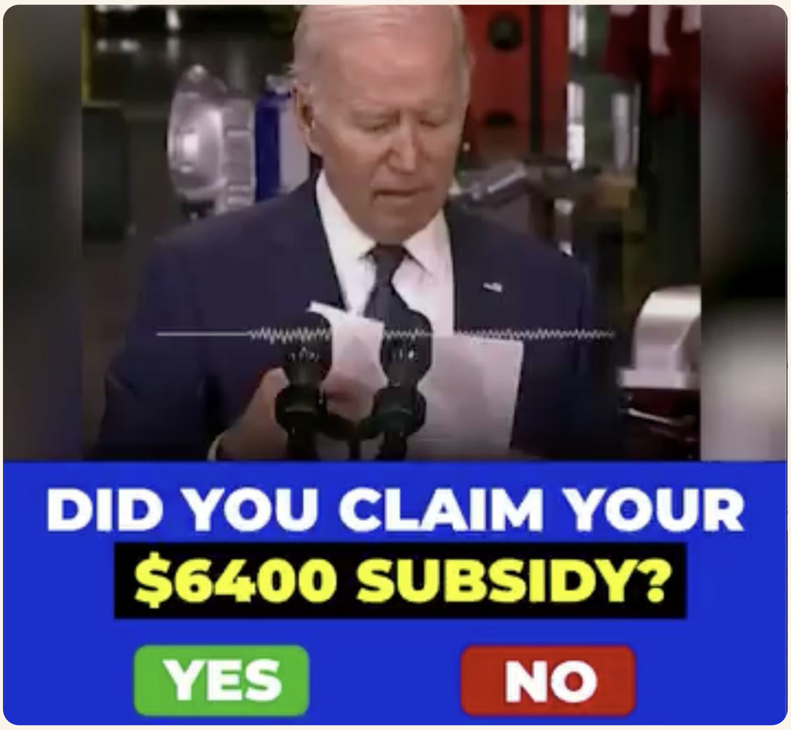
After the launch, scammers focus on amplifying the reach of their scam targeting similar population clusters to maximize the financial returns.

WATCHLIST

Scams

Deepfake Scams

See ↗

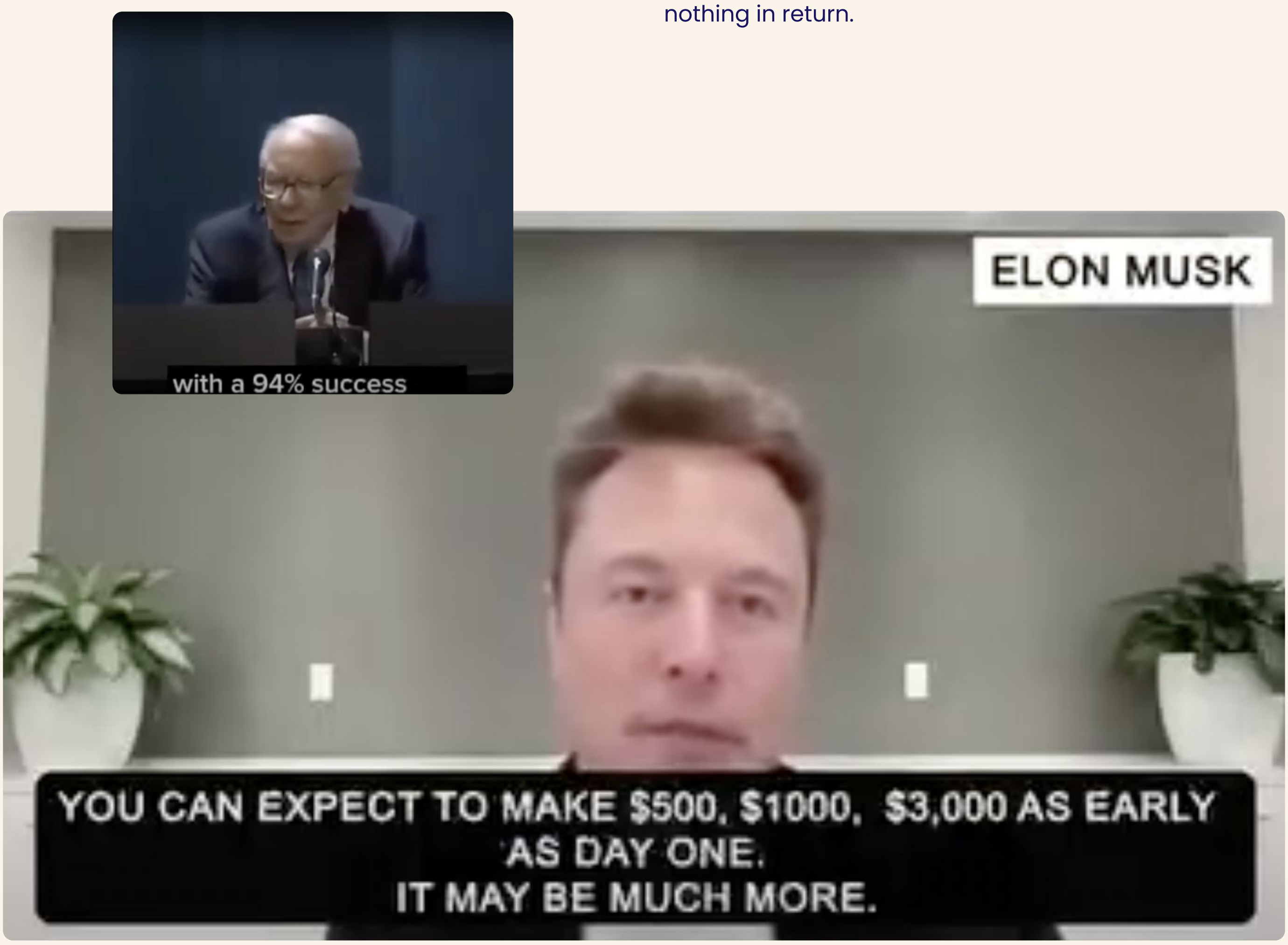


SPOTLIGHT

The Quantum AI Scam

The Quantum AI scam is a sophisticated online investment fraud that uses fake celebrity endorsements and deepfake technology to deceive people. This scam claims to use advanced technologies like AI and quantum computing to generate high profits for investors.

The scam typically involves the use of social media ads and fake news articles, exploiting a combination between Lip Sync deepfake algorithms and voice cloning in order to perfectly replicate well known public figures like Elon Musk claiming they endorse or have profited from the platform. These endorsements are entirely fictitious and designed to lure unsuspecting victims into investing money on platforms that offer nothing in return.


with a 94% success
ELON MUSK
YOU CAN EXPECT TO MAKE $500, $1000, $3,000 AS EARLY AS DAY ONE.
IT MAY BE MUCH MORE.

Watchlist

See ↗

SPOTLIGHT

The Quantum AI Scam

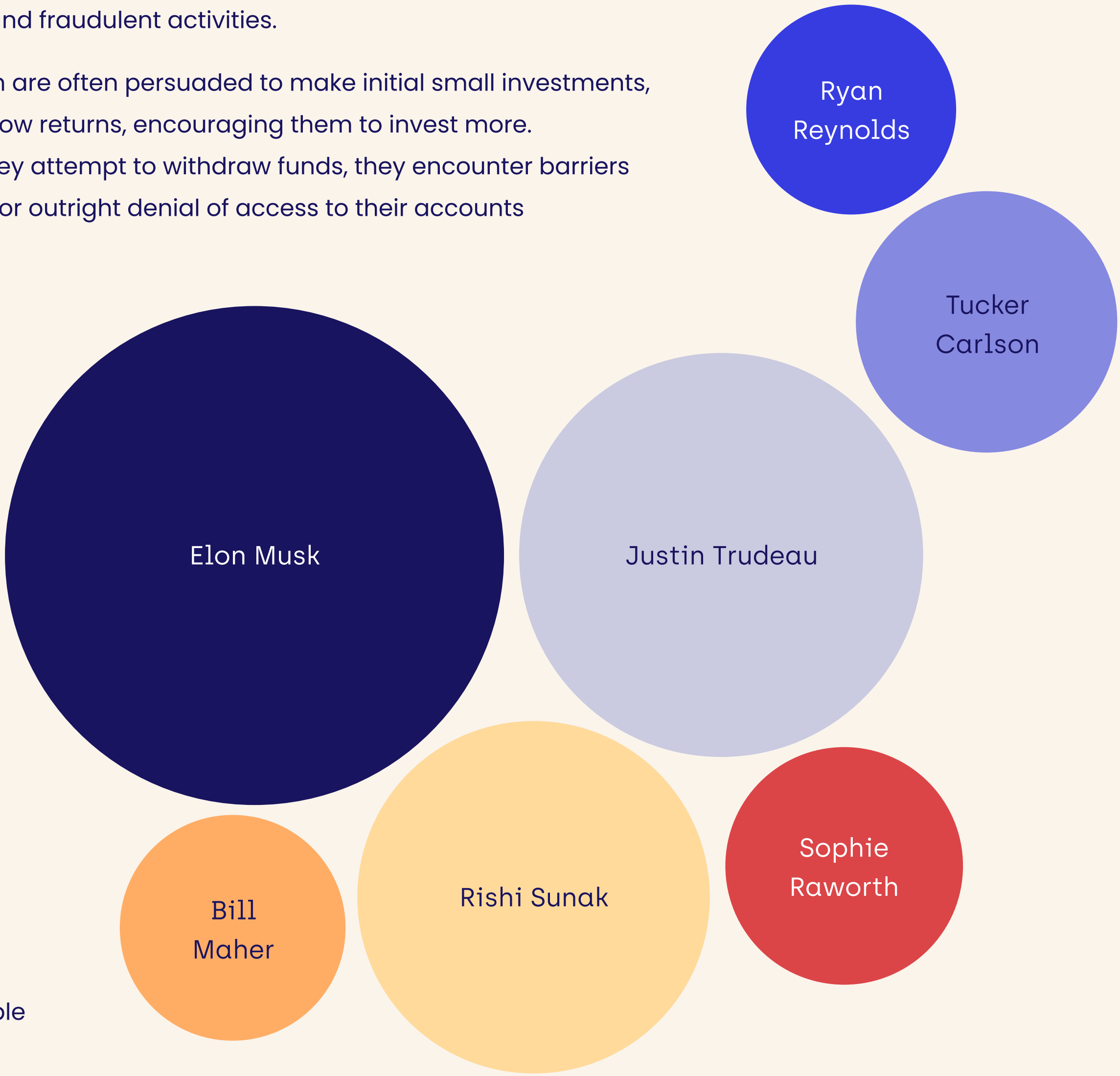
The complexity of this scam lies in how it's spread and made to seem more believable, tailored to each country. First, the scam is introduced to different areas, using Ad Network targeting tools. Then, fake local reporters start interacting with well-known global figures to make the scam seem legitimate. This tricks potential victims into signing up and making their first deposit by showing them fake data and results.

48

Countries heavily targeted by this scam, mainly in Asia, Europe and North America

The Quantum AI scam includes promises of guaranteed high returns with little or no effort, aggressive sales tactics, anonymity of the operators, and lack of proper documentation or licensing. Regulatory bodies in several countries have issued warnings against dealing with Quantum AI due to its unlicensed status and fraudulent activities.

Victims of the scam are often persuaded to make initial small investments, which appear to show returns, encouraging them to invest more. Eventually, when they attempt to withdraw funds, they encounter barriers like additional fees or outright denial of access to their accounts



Most targeted people on this scam:

KYC Frauds

KYC FRAUDS

Deepfakes in the online biometric identity verification systems, are now being leveraged in sophisticated real-time attacks that involve injecting manipulated facial imagery into live camera feeds.

Utilizing mobile emulators and virtual cameras, attackers can convincingly alter a video stream to bypass facial recognition and liveness checks during online KYC (Know Your Customer) onboarding processes. This is typically achieved through advanced software that enables face-swapping in real time.

The attackers, often part of organized cybercrime syndicates, possess in-depth knowledge of both artificial intelligence technologies and the security systems they aim to compromise. These tools are not just advanced in their technical capabilities but also increasingly accessible, making it easier for fraudsters to execute these operations seamlessly.

The reason behind such fraudulent activities is largely financial, targeting banks, online payment platforms, and any entity that relies on digital identity verification. The prevalence of such tactics points to significant vulnerabilities within current security measures, highlighting a growing need for robust countermeasures against these real-time deepfake technologies.

CYBER CRIMINALS

FRAUDSTERS

What, Why. Where, Who

WEB AND MOBILE APPS REQUIRING IDENTITY VERIFICATION

WHAT

Deepfake KYC fraud involves the use of sophisticated AI-generated imagery or manipulated video to bypass biometric verification processes used in online banking, financial services, and other web/mobile platforms.

WHY

The primary goal of these frauds is to steal face and voice biometrics in order to bypass KYC checks on web and mobile apps, with the intent to deposit and launder money coming from illicit online activities.

WHERE

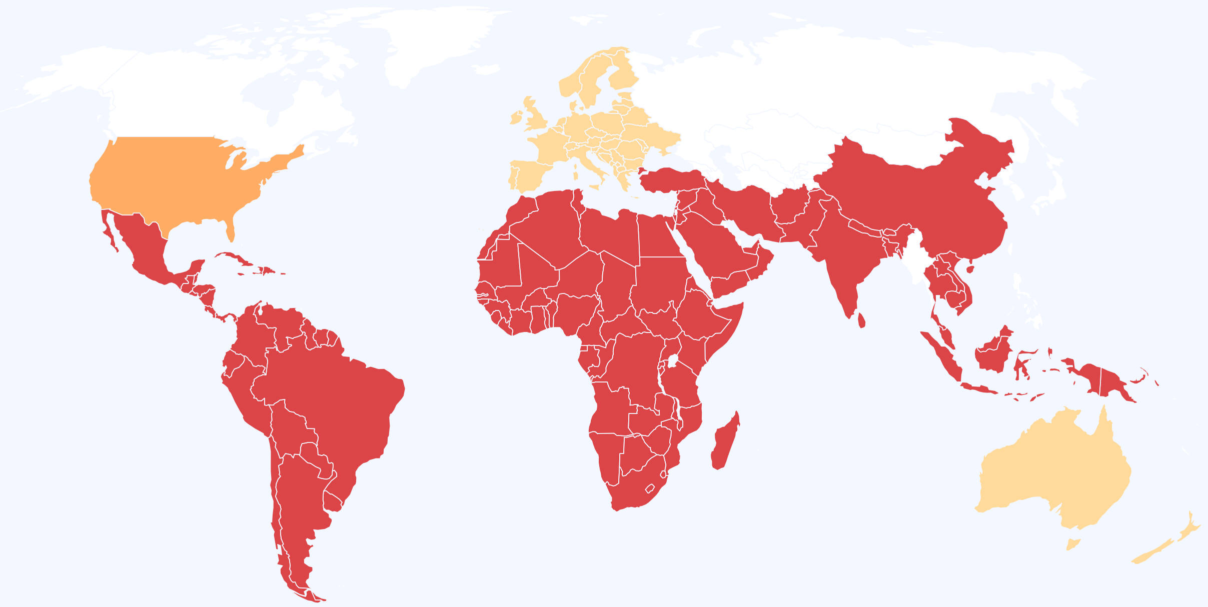
Deepfake KYC frauds occur primarily in online platforms that require biometric authentication. This includes online banking, digital payment systems, and any services that use facial recognition for identity verification. As businesses increasingly rely on remote identity verification technologies, the potential for such frauds expands across financial sectors globally.

WHO

The perpetrators of these frauds are often skilled fraudsters and cybercriminals with access to advanced AI and machine learning technologies. These individuals or groups are technologically savvy and possess the necessary resources to create and deploy high-quality deepfakes capable of deceiving biometric systems.

Most attacked countries

The map below shows the percentage of generative AI techniques used for online KYC systems' attacks by region.



🟥 Red

Extremely high number of deepfake attacks against online KYC systems

🟧 Orange

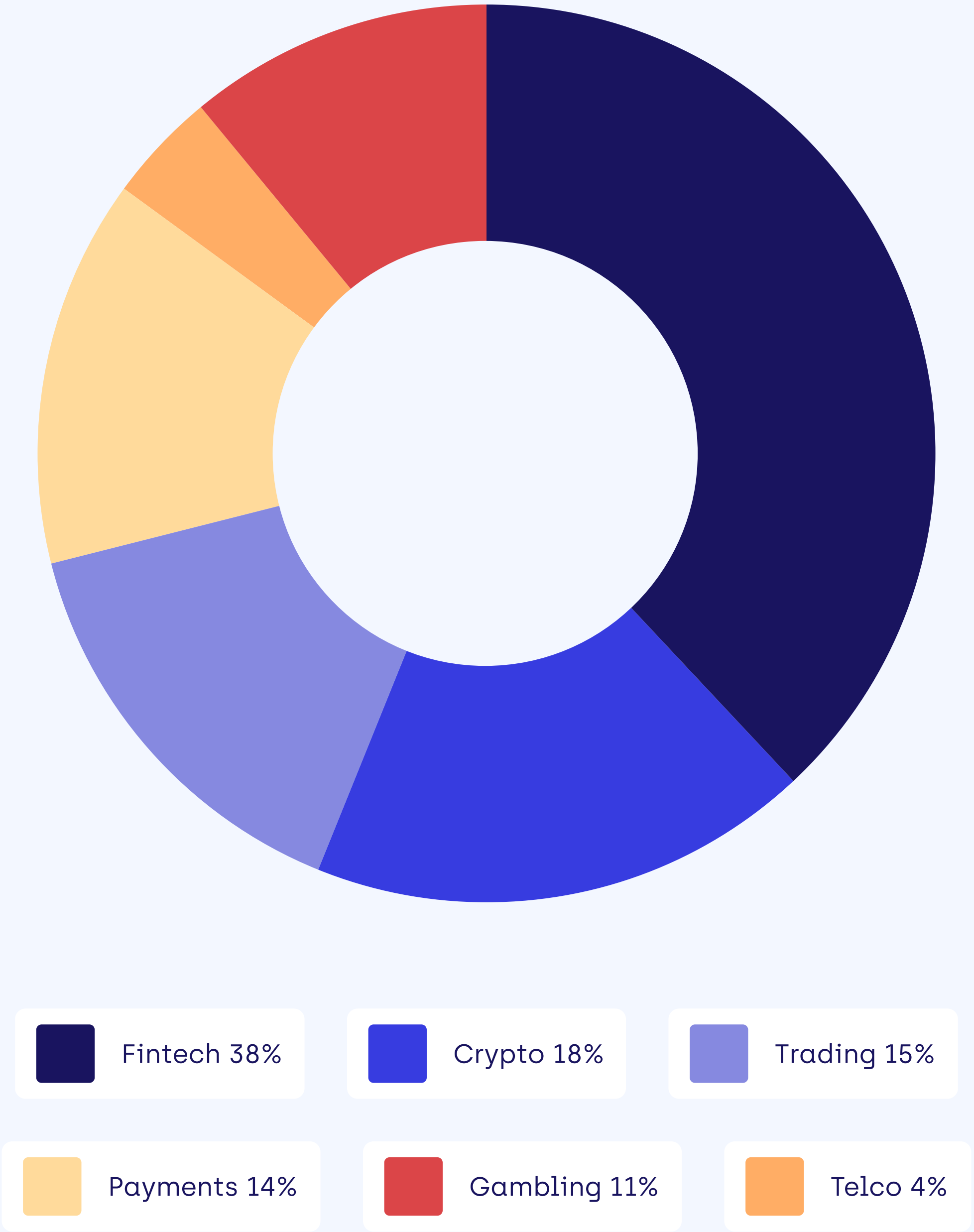
High number of deepfake attacks against online KYC systems

🟨 Yellow

Limited number of deepfake attacks against online KYC systems

Most attacked sectors

The chart shows deepfake KYC frauds most attacked sectors



Anatomy of deepfake KYC attacks

Passport Gathering

The bad actor collects personal information and photographic IDs of potential victims. This information can be sourced through social engineering, data breaches, or dark web marketplaces.

Vcam and mobile emulator

Virtual cameras or emulators are software tools that allow the attacker to manipulate digital images and videos in real-time and feed this altered visual data to the application as if it were coming from a legitimate camera or mobile device.

Victim selfie download

This photo is critical as it forms the basis of the deepfake; it should be high quality and contain multiple angles of the victim's face to increase the believability of the fake identity.

Arming real time deepfake

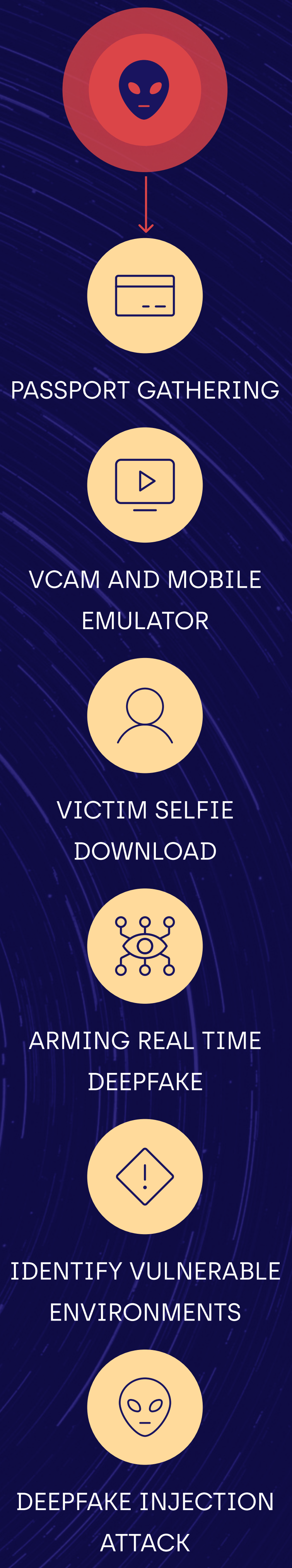
The actor creates a model that can generate real-time facial expressions and movements that mimic the victim, based on the input from the selected victim photo.

Identify vulnerable environments

This involves finding weaknesses in the application's security measures, such as outdated facial recognition algorithms or insufficient liveness detection capabilities, that can be exploited using a deepfake.

Deepfake injection attack

The prepared deepfake is deployed during the KYC process. The virtual camera or emulator feeds the generated images or videos directly into the application's biometric verification system. If the attack is successful, the system is deceived into recognizing the deepfake as the legitimate user, thus granting unauthorized access or approval.



PASSPORT GATHERING

VCAM AND MOBILE EMULATOR

VICTIM SELFIE DOWNLOAD

ARMING REAL TIME DEEPFAKE

IDENTIFY VULNERABLE ENVIRONMENTS

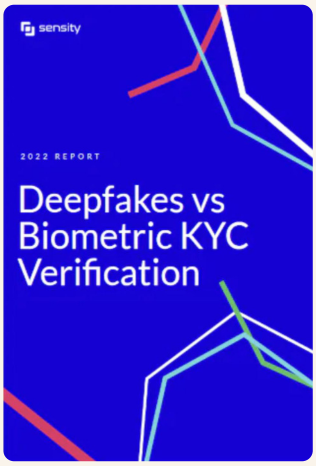
DEEPFAKE INJECTION ATTACK

REPORT

Read our previous research on deepfake KYC attacks

Deepfakes vs Biometric KYC Verification

Read ↗


sensity

2022 REPORT

Deepfakes vs Biometric KYC Verification

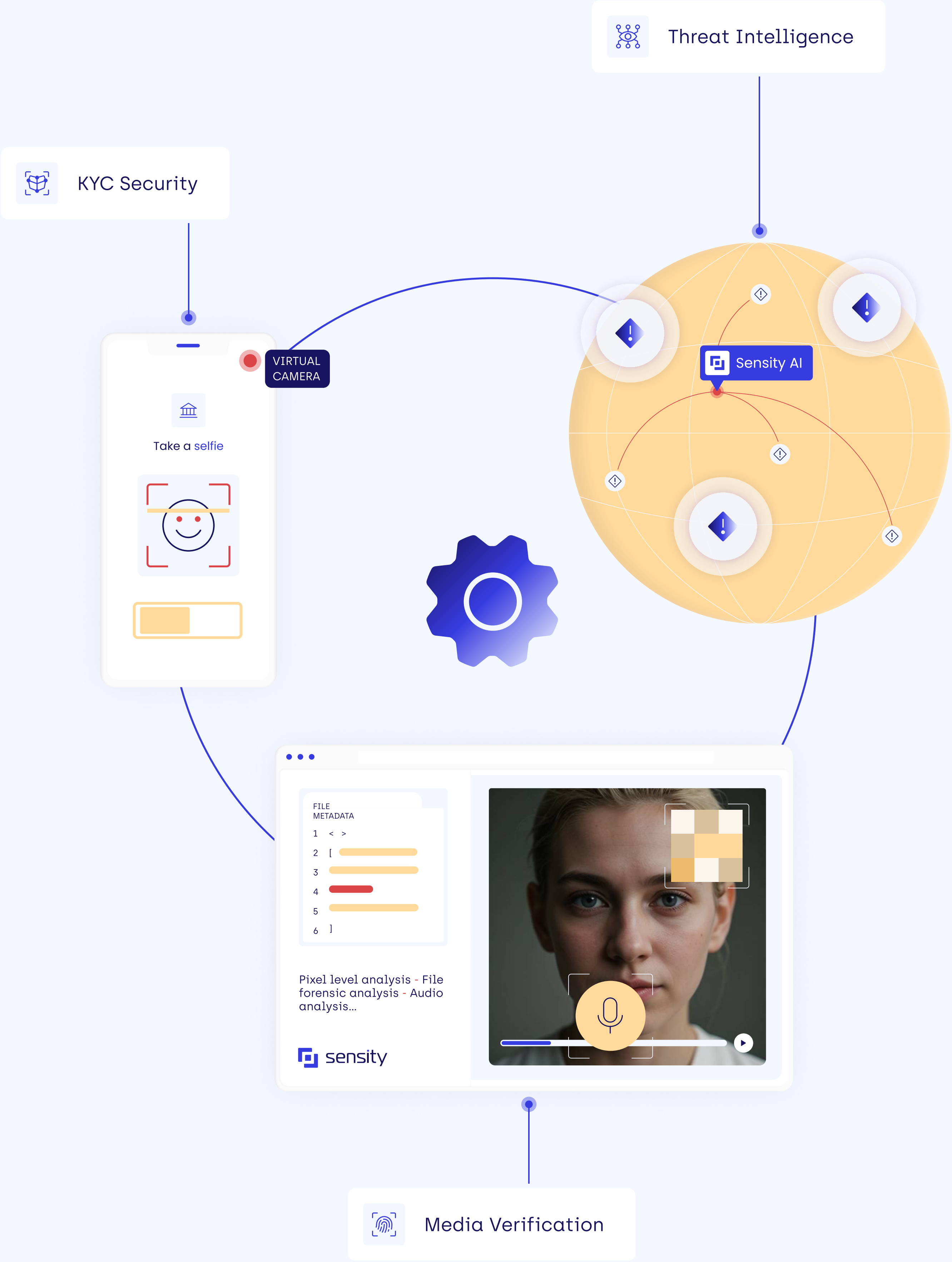
Next Challenges

As we look ahead, the landscape of deepfake technology presents a series of formidable challenges that demand our attention and action.

Addressing these issues will require concerted efforts from policymakers, technology developers, and the public to safeguard against the potential misuse of deepfakes and ensure the responsible development of AI technologies.

- AI-generated listings on travel marketplaces could erode consumer trust and disrupt the travel industry.
- The manipulation of insurance claims by AI could lead to significant financial losses and complicate the claims process.
- Deepfake romance scams pose severe emotional and financial risks, exploiting vulnerable individuals.
- Impersonation during video calls threatens the integrity of remote communications, while real-time voice cloning for phone scams introduces a new level of sophistication to fraudulent activities.

Sensity AI
All-In-One deepfake detection



Threat Intelligence

KYC Security

Media Verification

Sensity AI All-In-One deepfake detection

Media Verification

We equip companies and government agencies with the most advanced multilayer detection technology for AI-Manipulated and fully synthesized video, images and audio. Our pixel level analysis focuses on the content to spot signs of manipulation and synthesis such as: face swap, lip sync, face reenactment, face morphing and AI-powered human avatars. In addition we analyze the file structure in order to track from which device the file comes, which software has edited it and on what social media platform has circulated. Our voice analysis technology can assess whether voices into videos or audio files are AI Generated or not.

KYC Security

We provide web and mobile platforms with the best in class for preventing deepfake injection during online identity verification. We developed a modular solution that enables companies to build a customized detection suite based on the type of threat to solve. Sensity's SDK offers a strong countermeasure in preventing malicious actors deploying virtual cameras and mobile emulators. Our detection endpoint can assess in a range between 1 and 5 seconds whether a single frame liveness check or an active liveness check are being attacked using real time deepfake technologies.

Threat Intelligence

We adopt a threat intelligence approach to monitor and detect deepfakes on top of the AI-powered media verification suite. A dedicated threat analysis team, operates globally continuously new instances of deepfake content in the wild on every type of internet channel (including the darknet) thanks to our unique deepfake monitoring system built in 2019 that allows us to research deepfake content by geography, people, organizations and topics and stay up-to-dated about new type of threats. Over the past four years, this vigilant monitoring has enabled Sensity AI to amass the largest database of detected deepfakes. This additional security layer positions Sensity AI at the forefront of digital media verification and security.

sensity

The State Of Deepfakes 2024