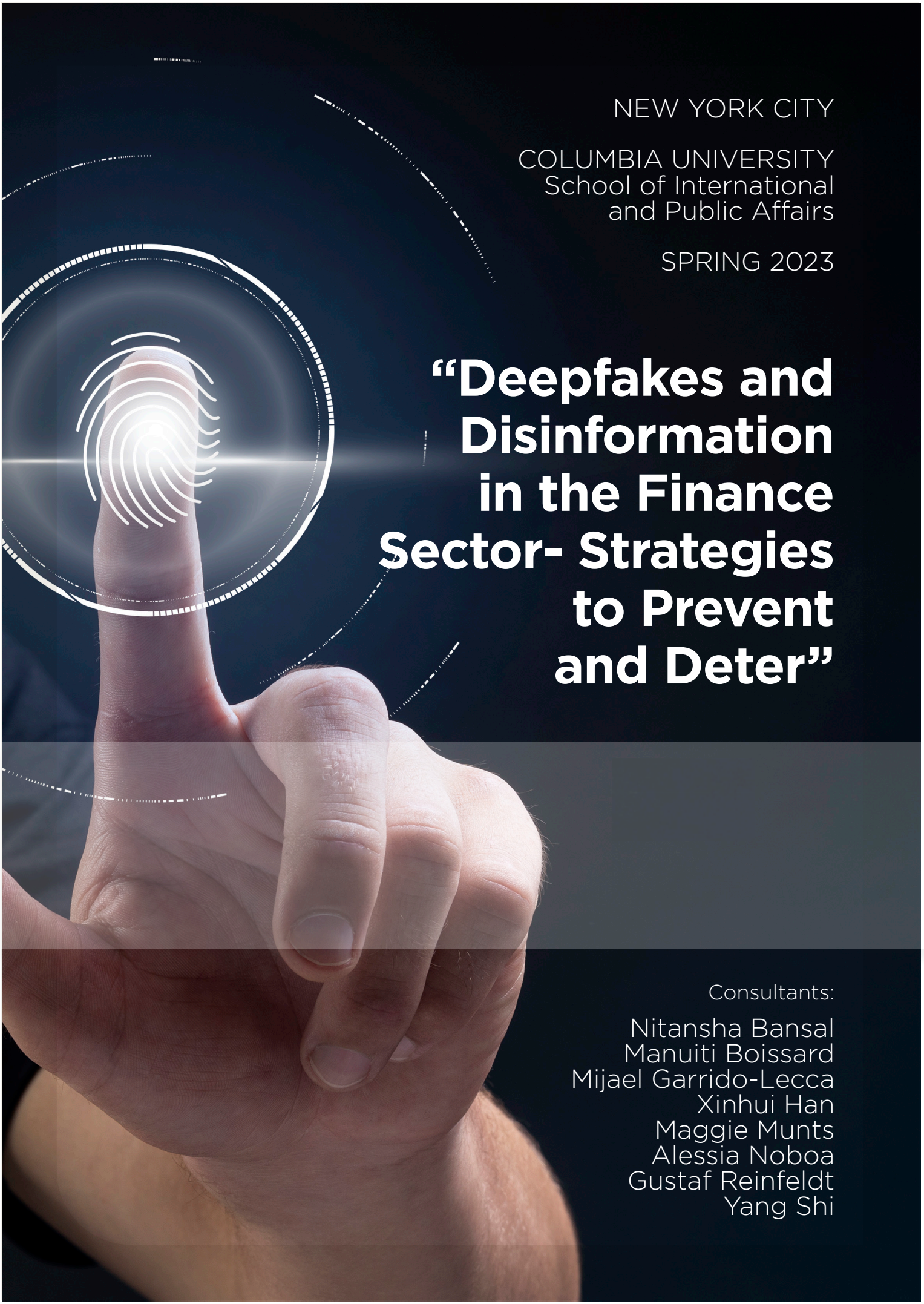


NEW YORK CITY

COLUMBIA UNIVERSITY
School of International
and Public Affairs

SPRING 2023

"Deepfakes and Disinformation in the Finance Sector- Strategies to Prevent and Deter"

Consultants:

Nitansha Bansal
Manuiti Boissard
Mijael Garrido-Lecca
Xinhui Han
Maggie Munts
Alessia Noboa
Gustaf Reinfeldt
Yang Shi

Deepfakes and Disinformation in the Finance Sector: Strategies to Prevent and Deter

COLUMBIA UNIVERSITY
School of International and Public Affairs

SPRING 2023

Contents

Contents	**2**
I) Executive Summary	**3**
II) Scope and Definitions	**4**
III) Impact of Disinformation and Deepfakes	**8**
IV) Focus: The Financial Sector	**10**
V) Solutions	**13**
Prevention	18
Detection	20
Response	22
VI) Case Study: Silicon Valley Bank	**22**
VII) Conclusions	**24**
Acknowledgements	**26**
References	**27**

I) Executive Summary

This report provides insights on the current state of deepfakes and disinformation in the financial sector. We found that while deepfakes have long been a tool for cyber threat actors, the recent commercial growth and expansion of generative AI technologies has resulted in deepfakes becoming notably more sophisticated, accessible, and easy to produce. This heightens the risks that threat actors will use deepfakes to impersonate executives and key personnel, degrade brands, defraud customers, influence markets, or run information operations. Disinformation risk has also escalated with the growth of these emerging technologies.

Our research methodology included a deep dive into deepfake and disinformation literature, case studies, and fifteen in depth interviews with private and public sector professionals. These interviews shed light on how deepfakes and disinformation have affected professionals across both the private and public sector, particularly those involved in work with the financial sector. Our team conducted this research as part of a consultancy between January and April 2023 in the context of the capstone requirement for our masters' degree program at Columbia University's School of International and Public Affairs (SIPA).

As new technologies complicate the cyber landscape, security professionals must update strategies to effectively prevent, deter, and respond to disinformation campaigns. Hyper-realistic deepfakes, undetectable to the human eye, pose a growing threat to the financial sector as they allow cybercriminals to outsmart even the most security-conscious employees.

By our analysis, increased employee training efforts, especially for high-profile executives susceptible to being 'deepfaked' or otherwise impersonated - from spear phishing campaigns to media presence - will be a crucial pillar for security controls to enhance organizations' resilience to deepfake and disinformation threats. Multi-factor authentication also becomes even more essential when seeing or hearing (in the case of audio deepfakes) is no longer a guarantee of authenticity.

This report will focus on defining key terms and examines the techniques, tactics, and tools that drive deepfakes and disinformation. It analyzes the specific impacts on financial institutions and recommends mitigation strategies for detection, prevention, and response. It also explores how new technologies arising from AI innovations can be leveraged for a variety of purposes across all sectors.

In the following section, we define key terms in this report. Then we provide an overview of how deepfakes and disinformation campaigns can impact brands, operations, information security, and customer trust. We then take a closer look at impacts of deepfakes and disinformation to the current and future state of the financial sector. The report then proceeds with an overview of solution types for deepfakes and disinformation: prevention, detection, and response. We then examine Silicon Valley Bank as a case study of what can happen when content about a financial

institution circulates rapidly online. We conclude with a look at existing solutions for prevention, detection, and incident response that may be relevant to the financial sector's cyber defenses.

II) Scope and Definitions

Disinformation in this report will be defined as false or misleading information with the intent to deceive or mislead people or institutions. Disinformation is not a new concept for the financial industry, but the prevalence of contemporary information technologies and growing digitization of the financial sector,[1] have broadened the attack surface for disinformation campaigns. New, highly scaled technologies, like generative artificial intelligence, also introduce new security challenges and amplify existing threats.[2] While other kinds of inaccurate and harmful information (e.g., misinformation and malformation) are also areas of concern for the financial industry and global private sector, this report focuses specifically on disinformation and deepfakes as an emergent tactic and tool of disinformation.

In this report, the term ***deepfake*** refers to digital content that is manipulated or generated using artificial intelligence, large language models, or deep learning algorithms. Deepfakes tend to take the form of hyper-realistic videos, audios, or other forms of digital content that are actually fake.[3] With a deepfake it is very easy to make it look like someone is saying or doing something they never said or did.

Dark PR also continues to be a threat to corporate entities. Dark PR refers to unethical or deceptive practices that some public relations professionals or agencies may engage in to manipulate public opinion or to damage the reputation of an individual or organization. This could include spreading false or misleading information, using underhanded tactics to generate media coverage, or engaging in smear campaigns against competitors to result in losses.

There are several different ways to create a deepfake. Historically, one of the more common ways to create a deepfake involved using an autoencoder program to employ a face-swapping technique to generate a hyper-realistic image or video. It is largely used to deepfake people, not things. The process involves combining "target media" with a collection of real media of the person that the deepfake is targeting. The autoencoder program studies these media clips (image, video, or audio) to successfully mimic what the person looks like from different angles or conditions. It is then able to reformat the target image, video or audio to resemble the person accurately, thereby creating deepfake content.[4]

---

[1] Valenti, Jonathan, and Ryan Alderman. "Will the Shift to Digitalization in Banking Stick?" Deloitte Insights, Deloitte, 20 June 2022

[2] Bateman, Jon. "Deepfakes and Synthetic Media in the Financial System: Assessing Threat Scenarios." Carnegie Endowment for International Peace. July 2020. Web.

[3] Johnson, Dave. "What Is a Deepfake? Everything You Need to Know about the AI-powered Fake Media." Business Insider. Business Insider, 10 Aug. 2022. Web.

[4] Ibid

Generative Adversarial Networks (GANs) are another technology that has been used to improve and perfect a deepfakes, resulting in hyper-realistic image, video and audio. Deepfake models are created by training a computer with GANs to analyze, detect, and correct imperfections based on a highly detailed comparison of the real content to the deepfaked content. This is sometimes referred to as “training” the deepfake model. The more content of someone that exists and is publicly available, the easier it is to train a deepfake on their likeness.

Historically the generative technologies and technical expertise required to create deepfakes was consolidated in the hands of a relatively few academics and bad actors. But now, the generative technologies behind deepfakes have become widely accessible across the internet. Known as generative AI, these widely accessible models and tools from Open AI’s DALLE and ChatGPT to Midjourney and Re-speecher among hundreds of others, can create deepfakes and other kinds of synthetic media with minimal text prompting.

Generative AI refers to the artificial intelligence and large language models (LLM) that generate deepfakes or other hyper-realistic outputs with minimal human prompting. This includes synthetic image, audio, video, and text-based responses (e.g., Chat GPT), or even code. Trained on vast amounts of audiovisual data, generative AI outputs can mimic a person's voice, mannerisms, and facial expressions with remarkable accuracy by synthesizing vast amounts of audiovisual data. Within the last 12 months generative AI technologies have become markedly more accessible, efficient and realistic. Within the last 4 months, that trend has only accelerated with commercial generative AI providers releasing new and updated consumer-facing tools like Open AI’s ChatGPT 4.[5]

As generative AI continues to advance, and investment dollars continue to flow,[6] so does the risk of creating convincing deepfakes. Efforts to measure deepfake count online are becoming increasingly futile. Experts estimate that up to 90% of online content will be synthetic by 2025.[7] One of the last meaningful efforts to attempt to measure the quantity of deepfakes online was conducted in 2019, by Sensity AI (formerly known as Deeptrace). Because deepfakes were still fairly nascent, researchers were able to use keyword search on media websites to find over 15,000 deepfake videos.[8] Sensity AI found that by 2020, that number had doubled. They also found that 96% of deepfakes at that time featured non-consensual sexual imagery.

Now with generative technologies more widely accessible than ever before, deepfakes and other forms of synthetic media are significantly more present across the internet and are used by a wide range of threat actors to further deceptive aims.[9]

---

[5] GPT-4, https://openai.com/product/gpt-4.
[6] Wiggers, Kyle. “VCS Continue to Pour Dollars into Generative AI.” TechCrunch, 28 Mar. 2023
[7] Bandara, Pesala. “90% of Online Content Could Be Generated by AI by 2025, Expert Says.” 90% of Online Content Could be Generated by AI by 2025, Expert Says | PetaPixel, January 17, 2023. https://petapixel.com/2023/01/17/90-of-online-content-could-be-generated-by-ai-by-2025-expert-says/.
[8] Ajder, Henry, et al. “The State of Deepfakes.” DarkTrace
[9] Paul, Olympia A. "Deepfakes Generated by Generative Adversarial Networks." 2021. Web.

Generative AI and machine learning developments have allowed for new language models to increasingly spread disinformation. Content-generating chatbots like ChatGPT or Google Bard enable cybercriminals to easily produce and refine highly believable content quickly and at scale.[10] Moreover, these generative chatbots have demonstrated a tendency to "hallucinate" and produce answers and text that is "confidently wrong," generating misinformation with authority.[11] These content generators allow anyone to create large swaths of distinct, persuasive language quickly - whether or not its contents are actually accurate. This accelerates the risk that they will be used by threat actors to drive disinformation campaigns or used by unknowing consumers to unintentionally create and propagate misinformation.[12,13]

Because generative AI models produce hyper-realistic outputs, deepfakes and other forms of synthetic media are by nature, difficult for the human eye (or ear) to detect with high accuracy. This makes them a particularly dangerous technology when we consider imitation and impersonation as common fraudulent behaviors. Without clear, resilient authentication protocols, threat actors hiding behind synthetic content (voice, video, live deepfakes, text messages, etc) may be harder to identify.

The content produced will be able to influence people not solely because of the high quality of the content but because of the amount of time it may travel undetected within a digital ecosystem before being detected. [14] Investigators will likely have to rely more heavily on other behavioral indicators to identify threats.[15]

Easy access to technologies like Re-speecher, Murf, Speechify and Listnr, have made vishing using audio deepfakes an increasingly popular tactic among threat actors.[16] Vishing allows malicious actors to impersonate their target easily and bypass voice authentication technology. This is a particular threat to the financial sector, as many banks use voice authentication to try to reduce fraudulent transactions and many customer interactions happen over the phone.[17] Live deepfakes are also becoming increasingly sophisticated[18] meaning that it may be easier for threat actors to imitate customer or employee voices in real time. Vishing using deepfake audio make

[10] Goldstein , Josh, et al. "Generative AI Is Enabling Fraud and Misinformation - Here Is What You Should Know." Center for Security and Emerging Technology, 7 Mar. 2023,
[11] https://www.nytimes.com/2023/03/29/technology/ai-chatbots-hallucinations.html
[12] https://arxiv.org/pdf/2301.04246.pdf
[13] Goldstein , Josh, et al. "Generative AI Is Enabling Fraud and Misinformation - Here Is What You Should Know." Center for Security and Emerging Technology, 7 Mar. 2023
[14] Ibid
[15] François, Camille. "Actors, Behaviors, Content: A Disinformation ABC." Annenberg Public Policy Center, September 20, 2019. https://cdn.annenbergpublicpolicycenter.org/wp-content/uploads/2020/06/ABC_Framework_TWG_Francois_Sept_2019.pdf.
[16] Noone, Greg. "Listen Carefully: The Growing Threat of Audio Deepfake Scams." Tech Monitor, February 4, 2021. https://techmonitor.ai/technology/cybersecurity/growing-threat-audio-deepfake-scams.
[17] Ghosh, Soumik, and Ron Ross. "Deepfakes, Voice Impersonators Used in Vishing-as-a-service." Bank Information Security. 3 Dec. 2021. Web.
[18]Penner, Derrick. "TED Vancouver: Live Video Switch Illustrates Chilling, Thrilling Potential of Artificial Intelligence." Vancouver Sun, April 18, 2023. https://vancouversun.com/news/local-news/ted-live-video-switch-illustrates-chilling-thrilling-potential-of-artificial-intelligence.

people extremely vulnerable to being deceived via phone call, which was the case for a UK CEO in 2019 who transferred $243,000 after believing he had been on the phone with his German colleague and did not realize it was a fraudster until it was too late, and the money had already been transferred. [19]

Similar Generative AI technologies that mimic human text communications can also accelerate smishing threats. This type of deepfake can be essential for criminals using social engineering. As cybercriminals can make emails or other communications look legitimate, they can readily con consumers or target professionals in business email compromise (BEC) attacks.

A note on ethics:

Deepfakes have the ability to portray people and situations with a high degree of realism. According to recent research, humans are no longer able to distinguish AI-generated faces from real human faces.[20] This raises obvious concerns around disinformation and non-consensual distribution. Now as generative AI becomes increasingly available and easy to use, natural questions arise about the ethics of using this kind of hyper-realistic technology to generate more innocuous kinds of content, like marketing materials, employee training videos, or imaginative virtual worlds. Many generative AI companies, like Synthesia, have positioned themselves as B2B solutions providers for these kinds of use cases.

Two clear best practices have emerged for creating and distributing synthetic media: 1) disclosure and 2) consent. The Partnership on AI's Responsible Practices Framework for Synthetic Media, launched in Feb 2023, emphasizes both disclosure and consent as best practices when creating, sharing, or hosting synthetic content.[21]

- **Consent:** Prior notification and permission must be obtained from individuals whose information or image is being used for deepfakes.

- **Disclosure (Transparency):** Maintaining transparency is crucial, and individuals have the right to know if they are interacting with a program and if what they see is genuine. Companies also need to disclose the essential information to the public.[22]

With deepfakes, especially in the US, there is clear tension between 1st amendment protections around speech, and other legal protections around likeness, abuse, non-consensual distribution. In our research we also found that there is ample debate as to whether any areas should be

[19] Damiani, Jesse. "A Voice Deepfake Was Used to Scam a CEO out of $243,000." Forbes. Forbes Magazine, 03 Sept. 2019. Web.
[20] Farid, Nightingale SJ. "Ai-Synthesized Faces Are Indistinguishable from Real Faces and More Trustworthy." Proceedings of the National Academy of Sciences of the United States of America, U.S. National Library of Medicine
[21] Responsible Practices for Synthetic Media - Partnership on Ai. https://partnershiponai.org/wp-content/uploads/2023/02/PAI_synthetic_media_framework.pdf.
[22] Hongo, Hudson. "Pai Seeks Public Comment on the Synthetic Media Code of Conduct." Partnership on AI. 02 Nov. 2022. Web.

entirely off-limits for deepfake use, where the risk of misleading viewers far outweighs the convenience of AI-generation. Political campaigns are one such area where some regulators have started to take steps to clarify deepfake policies. In March 2023, the state legislature in Washington state introduced a bill that would provide political candidates with legal recourse should their likeness appear in a deepfake.[23] However, given that political satire is protected speech within the US, there is ample debate as to how to mitigate deepfake risk without impinging on first amendment rights. It is clear however, that greater accountability is needed, across developers, creators, distributors, and hosts.

III) Impact of Disinformation and Deepfakes

In March of 2022, a deepfake of Ukrainian president, Volodymyr Zelenskyy circulated on social media. In the video, a fake, but uncanny Volodymyr Zelenskyy instructed Ukrainian troops to stand down. The video was later removed as platform officials at Facebook, Youtube, and Twitter cited it as a violation of their platform policies.[24] A year later, researchers at Graphika found that commercial deepfake technology from B2B generative AI provider Synthesia had been used in a Chinese influence operation.[25]

As generative technologies become increasingly accessible, deepfakes pose growing risks to information integrity, geopolitical stability, and democracy around the world. As regulators worldwide grapple with if and how to regulate deepfake creation and distribution, the private sector will need to assess and mitigate its own exposure to deepfakes and associated risks.

However, as deepfakes approach higher and higher levels of realism, and the technologies to create them become more accessible and efficient, the private sector will have to pay special attention to the evolving threats to executives, brands, business operations, information security, and customer trust.

- **Threat to Executives and Employees:** Because deepfakes can be used to easily impersonate people, they can pose a significant risk to company's executives and, by extension, the business verticals they oversee. It is also easier to 'train' a model on someone's likeness if there is more public imagery, video, and audio of them. This means that higher profile individuals within a given company or institution are often at higher risk of being deepfaked and having their likeness abused. There is a risk that unsavory deepfaked content of key personnel could be used to extort sensitive information.

[23] Pereira, Ivan. "Washington State Bill Would Provide Safeguards against 'Deepfake' Political Ads." ABC News, ABC News Network

[24] Allyn, Bobby. "Deepfake Video of Zelenskyy Could Be 'Tip of the Iceberg' in Info War, Experts Warn." NPR, NPR, 17 Mar. 2022

[25] Satariano, Adam, and Paul Mozur. "The People Onscreen Are Fake. The Disinformation Is Real." The New York Times. The New York Times, 07 Feb. 2023. Web.

- **Threat to Brands:** Because deepfakes can be a powerful visual tool to misrepresent or mislead, brands face unique risk. While many generative AI platforms have specific guardrails to prevent brand integrity abuses, such guardrails do not protect against the full range of threats brands may face from generated content. For example, certain brands are closely affiliated with more general imagery and key words (e.g., a box of tissues = Kleenex) that could easily be generated with new generative AI platforms. Furthermore, if a large language model like ChatGPT hallucinates misinformation about a given brand or individual, users may take that information as fact without verifying, which could have negative consequences for brands and individuals alike.

- **Threat to Operations:** As an increasingly sophisticated and easy to access tool for threat actors, there is also a risk that deepfakes will be used to disrupt day-to-day business operations. For example, a deepfake of a CEO could be used to deceive company employees into believing that there is an impromptu company-wide holiday.

- **Threat to Information Security:** As discussed earlier in this report, deepfake content, especially audio and imagery, can become a tool for more sophisticated phishing, vishing, and smishing campaigns. If used to circumvent or trick authentication protocol, deepfakes could help threat actors gain unwanted access to important data, systems, or personnel.

- **Threat to Customer Trust:** As outlined above in the “note on ethics," there is also risk in using deepfake technologies without taking appropriate measures around disclosure and consent. If a company or institution uses deepfakes for commercial purposes and does so without appropriate disclosure and consent, it could severely damage customer trust. Building and maintaining trust with customers is essential to the success of any business or institution, and deepfakes can undermine these efforts by sowing seeds of doubt and uncertainty.

It is crucial for companies to be proactive and prepared to mitigate circulating deepfakes or other forms of disinformation. Having a plan of action can mean the difference between effectively managing the situation or facing chaos. In a survey conducted by TechInformed, research shows that 80% of businesses acknowledge the negative effects of mis and disinformation in all sectors, with 54% of business leaders expressing concern that these effects will only worsen over the next 20 years.[26] A recent report from Gartner, predicted that by 2022, 80% of marketing teams will develop content authenticity functions by 2027 to combat synthetic media, and mis and disinformation risk.[27] As shown by these statistics, it is essential for companies to prioritize the development of effective strategies and protocols to combat the spread of disinformation.

[26] Bishop, Adrian. "How Misinformation Can Impact Businesses." TechInformed. 06 Sept. 2022. Web.
[27] Gartner. “Gartner Predictions for CMOs Show AI, Social Toxicity, and Data Privacy Forge the Future of Marketing,” April 17, 2023. https://www.gartner.com/en/newsroom/press-releases/2022-12-13-gartner-predictions-for-cmos-show-ai-social-toxicity-and-data-privacy-forge-the-future-of-marketing.

IV) Focus: The Financial Sector

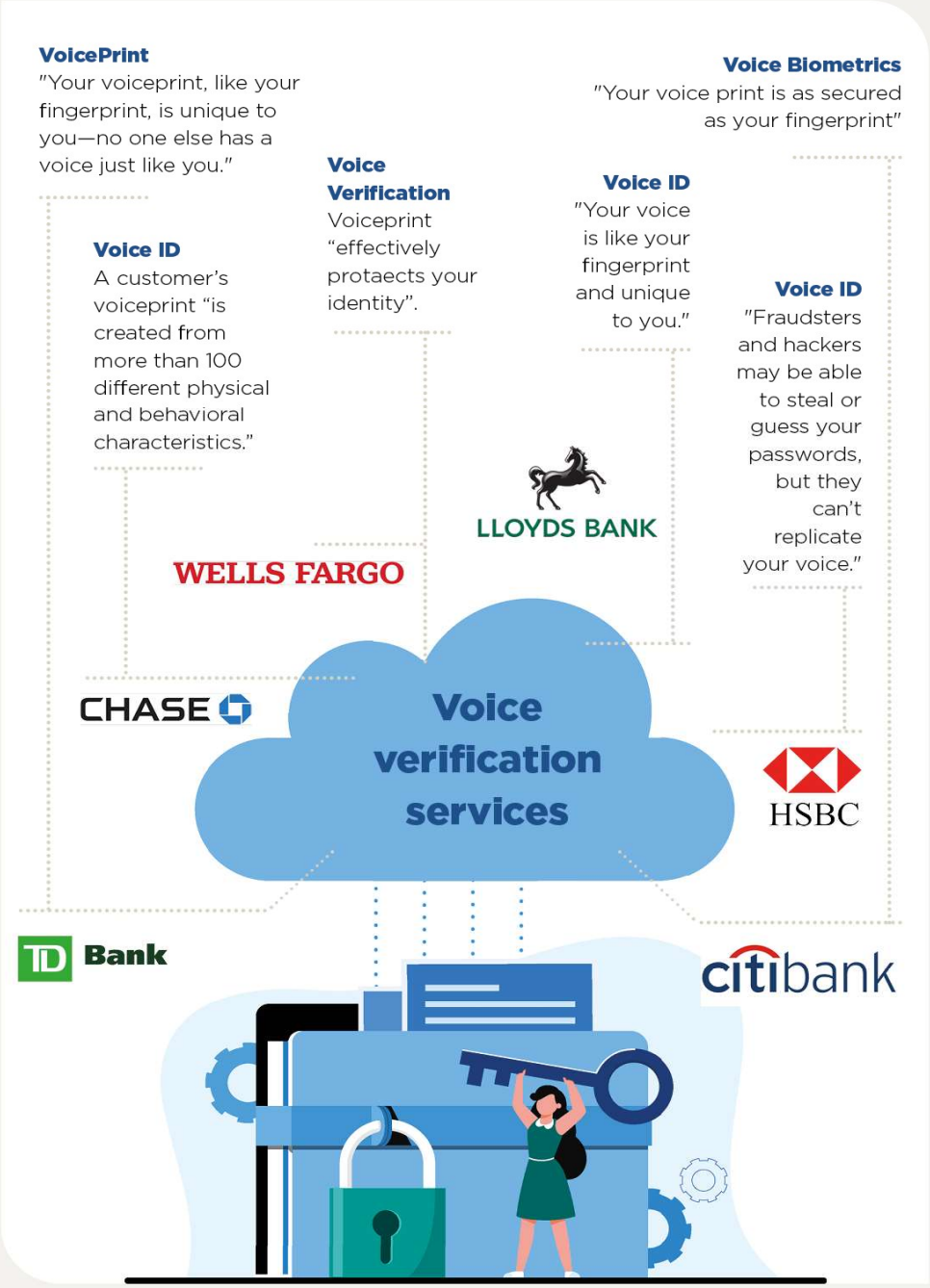
The financial industry has long been a leader in digitization, and an early adopter of digital processes and offerings for customers. Some experts estimate that financial services account for as much as 20-25% of the global economy.[28] The financial industry is a sizable, material, digital target for cyber threat actors, particularly cybercriminals. Financial institutions also have significant regulatory and reporting responsibilities given their role in the economy at large. Given the sizable cyber risk and regulatory stipulations characteristic of the financial industry, financial institutions have long been at the forefront of cybersecurity best practices. They also face some of the most savvy and creative cyber threat actors.

Our research for this project focused on deepfake and disinformation threats to the financial sector. Some of the key risk vectors and areas of vulnerability that we found in our research that are especially salient for the financial sector include:

- **Reputational Risks:** Reputational losses can be massive for institutions that are hit by deepfakes and disinformation attacks. The sensitive customer data that is withheld by financial institutions is highly valuable. Reputations are highly vulnerable when an institution falls victim to a mis/disinformation or deepfake attack. With low barriers to entry, incentives run high for cyber criminals. Aside from audio or visual deepfakes, generative AI can generate texts using large language models, generate fake images including ID cards, generate fake customer records or false records of transactions, and automate fraud by creating multiple fake IDs. [29]

- **Vulnerability in Content-Based Authentication Protocols:** As banks increasingly rely on voice ID, fraud can become easier for cyber criminals that can generate audio deepfakes using new large language models. One AI voice generator, Speechify, has been downloaded on the app store over 20 million times.[30] As of March 2023, a total of 47 open-source voice cloning software programs can be found on GitHub, amplifying the availability and ease of access to such technology. [31] If customers become aware of the ease to bypass front line security measures, their trust in the bank may be eroded which can lead to further actions such as pulling their money out. [32]

[28] Economist Intelligence Unit. “Financial Services Industry Trends | Economist Intelligence Unit,” n.d. https://www.eiu.com/n/global-themes/financial-industry-hub/.
[29] Salomon, Sanjay. “Generative AI Kicks off the 'Awesom-O' Age of Fraud.” Feedzai, 1 Feb. 2023,
[30] Weitzman, Cliff. “Deepfake Voice.” Speechify, 2 Mar. 2023,
[31] "Build Software Better, Together: Voice Cloning." GitHub. 2023. Web.
[32] Bateman, Jon. “Deepfakes and Synthetic Media in the Financial System: Assessing Threat Scenarios.” Carnegie Endowment for International Peace, July 2020



33

[33] How I Broke Into a Bank Account With an AI-Generated Voice. "How I Broke Into a Bank Account With an AI-Generated Voice," n.d. https://www.vice.com/en/article/dy7axa/how-i-broke-into-a-bank-account-with-an-ai-generated-voice.

- **Market Manipulation:** Other forms in which deepfakes can affect banks and financial institutions are through stock manipulation. Audio or video deepfakes can be used to falsify an endorsement for a product or a company that can later lead to an issue in investor relations. Consumer mindsets can also be altered which can lead to a decrease in activity for the bank as customers can perceive an institution as untrustworthy. Disinformation has been proven to have a market effect as demonstrated by the Syrian Electronic Army's 2013 hack of the Associated Press which led to a movement of $136 billion. [34] The hacking of the Associated Press involved a tweet falsely claiming that a bomb had exploded at the White House and ultimately injured the president. [35] The speed of information spreading is crucial to the impact malicious actors can have and with the use of bots paired with a fake audio or video recording, rumors can appear to be true very easily.

- **Risk to Credit Ratings:** Credit ratings can also be affected by false narratives spreading. Banks need to be particularly vigilant about this because customers can be readily misled and in turn affect the standing of the bank. As fraud claims and fraud activity becomes more prolific within the financial sector, banks will have to pay special attention to these issues if they want to stay afloat. It is important to maintain cyber risk in mind when discussing credit ratings. [36]

- **Scaled and Accelerated Cybercrime:** One example of generative AI making cybercrime easier is ChatGPT. Large language models such as ChatGPT or Google Bard accelerate the access that attackers have to knowledge and technical code without them having to necessarily learn new skills. [37] Models like this also make business email compromise readily available as criminals are easily able to mock the tone of someone using ChatGPT or other large language models. [38]

Deepfakes are a newer, emergent threat. Collecting information from interviews and research on trends and best practices vis a vis deepfake threats can help an institution to prevent, detect, and respond to deepfake attempts by threat actors more adeptly. That said, we found in our research and expert interviews that there is not yet a clear, unified protocol for responding. There are, however, several approaches that can help to mitigate risk. The following section explores possible solutions and solutions providers.

[34] Fisher, Max. "Syrian Hackers Claim AP Hack That Tipped Stock Market by $136 Billion. Is It Terrorism?" The Washington Post, WP Company, 1 Dec. 2021,
[35] Ibid
[36] Ibid
[37] Harr, Patrick. "Generative AI Changes Everything We Know about Cyberattacks." Dark Reading, 23 Feb. 2023,
[38] Ibid

V) Solutions

There is no “silver-bullet” to eliminate the range of risks posed by deepfakes, but there are solutions that can help with prevention, detection, and response. Institutions need to evaluate what controls and mechanisms are appropriate for their relative levels of exposure.

As we explore solutions, it is important to note that different solutions may be appropriate for different cultural contexts. A study conducted by the University of Chicago showed that 95% of Americans were worried about disinformation.[39] In contrast, our interviews showed that disinformation is less of a concern in Sweden because of government investment in media literacy education and the high level of media literacy among Swedes.[40] Based on our interviews with Swedish experts in the financial field and cyber security, they were confident that media literacy was key to not fall victim to false narratives gaining popularity.

The three solution sectors we have focused on are prevention, detection, and response. In each sector we have included some vendors that work within each space, but this is by no means an exhaustive list of all the vendors, organizations, and individuals working to mitigate the risk posed by deepfakes and other forms of disinformation across sectors. Although each sector is expanding as technology evolves, we have found these providers and solutions to be relevant to financial institutions.

[39] "The American Public Views the Spread of Misinformation as a Major Problem - AP-NORC." AP NORC University of Chicago. 8 Oct. 2021. Web.
[40] Interview with Swedish government officials

PROVIDERS

Prevention

C2PA

Industry coalition that published the technical open standard on content provenance and authenticity. Steering committee åmembers include Microsoft, Adobe and Truepic. Heavy govt advocacy efforts.

Benefits

Champions of an industry-leading standard. Some heavy hitters in the content space have thrown their weight behind C2PA and it consolidated a lot of the disparate work that was going on in this space. They are open to new members joining the coalition, which could have PR upside in addition to the security benefits of using technical content provenance for content-driven processes.

Limitations

Just an industry coalition, not a technical implementer - for that you'd ultimately need to work with a company that's already C2PA compliant.

TRUEPIC

CERTIFIED. ORIGINAL.

Truepic builds secure, interoperable content authenticity and transparency tech that would let you implement the C2PA's open standard and secure content-driven processes and communications. They are primarily focused on securing photos and videos. They have several award-winning tools used by leading insurance companies, lenders, development orgs to secure their content pipelines and photo driven processes with secure photo and video capture.

Benefits

Well established in this space since 2015. C2PA compliant. Member of CAI. Heavy-hitting engineering team. Well connected with thought leaders like Hany Farid and Nina Schick who sit on their advisory board. Best in class prevention for fraud and securing digital content driven processes and communications at scale. Have a great track record of working with large enterprise clients

Limitations

Not an all in one solution, but really good at what they do within their vertical, especially for photo and video verification.

OARO

Offers similar services to Truepic, but not a member of C2PA and not C2PA compliant. Provides tools to authenticate visual media using blockchain.

Benefits

Attaches verified data to photos and videos at the moment of capture. Has worked with partners in financial services.

Limitations

Not C2PA compliant, limited to visual media. Limited information on interoperability.

Detection

Detection for Text/ Documents

inscribe

Automated detection of document manipulation/ synthetic documents.

Benefits

Advertise 50% reduction in application review time, $80m+ in fraud caught per month, 25x return on investment, 200+ hours work saved per week. Save time and resources by automating document parsing, classification, and data matching. Track record with enterprise clients.

Limitations

Solely focused on documents, would need to work with them + other providers for holistic solutions.

Detection for Audio

REVA

Detection tool for audio tampering, manipulation or deepfakes

Benefits

Provides real time alerts to support tracking and response. Focus on audio deepfakes is important for account protection.

Limitations

Solely focused on audio, would need to work with them + other providers for a more comprehensive solution.

Pindrop

Leader in voice authentication to prevent voice spoofing, deep fake audio, etc.

Benefits

Established track record in industry especially with banks. Offers live detection for anti-fraud.

Limitations

Solely focused on audio, would need to work with them and other providers for comprehensive solution.

Detection for Visual Conten

FakeCatcher

the world's first real-time deepfake detector

Offers real time deep fake video detection using Intel hardware and software. Specifically looks at blood flow in the pixels of a video and uses deep learning to detect if it is fake or real.

Benefits

Real time detection within milliseconds So far, they claim a 96% accuracy rate

Limitations

Just launched at the end of 2022, so hard to really measure full impact/effectiveness at current stage and scale. Unclear how it will fair as visuals become increasingly realistic and generative technologies continue to improve in real time.

An open access deepfake dataset provided by Google & JigSaw to help people developing detection tech train their detection models more effectively.

Benefits

Google backed, open access

Limitations

Ultimately just a data set for people to train detection technologies.

DUCK DUCK GOOSE

Deepfake detection startup that uses ML to identify manipulated videos and images.

Benefits

Used by the Dutch House of Representatives, and Netherlands Forensic Institute. Claims to detect deepfakes in images with an accuracy of 95%. Able to distinguish 3 types of deepfakes: FaceSwap, StyleGANs and Deep Puppetry

Limitations

Focuses primarily on human faces. Does not identify deepfake audio. Like any AI solution - needs to be adequately trained on newest deepfakes to achieve higher accuracy.

Detection

Detection for Multiple Content Types

sensity

Founded in 2018, Sensity is considered to be the world's first deepfake detection platform. Sensity AI offers a suite of services for liveness detection, face matching, ID documents, deepfake detection and others to authenticate images, videos, or PDFs.

Benefits

Works across multiple kinds of content to detect. Relatively more established than some other detection providers.

Limitations

Needs to be installed which could lead to other vulnerabilities.

PREVENCY®

A monitoring system to detect disinformation attacks early. Prevency also offers simulations and training to handle disinformation attacks and 'dark PR.'

Benefits

Brings PR into the disinformation conversation. Uses a combination of technological solutions and human resources to detect disinformation attacks. They also offer some training on best practices when it comes to response.

Limitations

Ultimately more PR than security focused. No published latency times for detection (product page currently only available in German). Not focused on other cyber risks like fraud. Cannot prevent an undisclosed deepfake from hitting the internet.

GCID
INTELLIGENT FORESIGHT

Indian startup offering digital forensic services for images, videos, and audio.

Benefits

In addition to forensics services, they also provide forensics training. Partnered with some larger cyber training organizations like KnowBe4.

Limitations

Not necessarily tailor made for deep fakes or the financial sector. May be better for after the fact attribution/ authentication. Generally more focused on cyber criminal acts like cyber terrorism.

Reality Defenders provides government-grade detection platform provides enterprises and entities of any size intuitive protection against damaging deepfakes and generative content.

Benefits

Reality Defender detects these deepfakes in milliseconds, with tools and models that detect against an ensemble of existing deepfake methods and potential methods to come.

Benefits

Reality Defender are great at detection, but one must keep in mind that they ptovide numeric "levels of confidence" that something is a deepfake. They also do not offer post-incident advisory services, so interpretation of these levels of confidence is up to the person seeing these numbers. The other issue is that Reality Defender uses AI against AI, and that creates certain level of uncertainty.

Response

alethea.ai
Trusted Synthetic Media

Benefits

Provides real-time alerts for disinformation that help detect patterns for companies using both technology and human analysts as resources. It has worked with financial institutions before. They focus on ethics within a largely unregulated environment.

Limitations

It is solely focused on disinformation campaigns and does not specialize on detection, prevention or response. It is also a start-up that has not been around for an extended period.

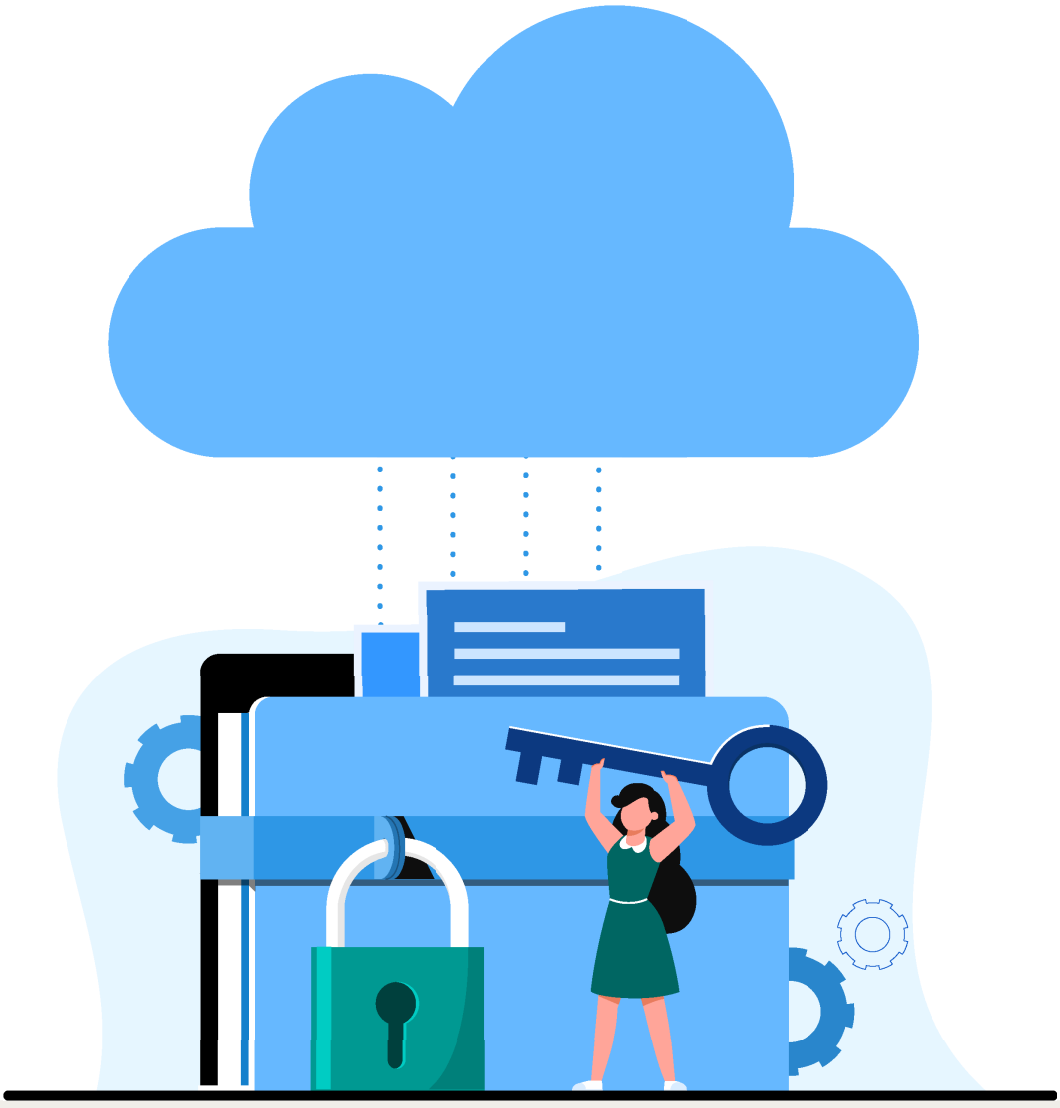
BLACKBIRD.AI

Benefits

Blackbird makes cutting-edge technology available to its clients; however, the wide range of clients that the company has allows the company to gain feedback from its own experiences and to make evidence harvested in an interdisciplinary process available to each client.

Limitations

As Blackbird is a relatively new company, the process of standardizing response protocols has just begun and they still depend a lot on the particular cases under discussion



Prevention

Preventing deepfakes and disinformation requires upstream interventions to authenticate users, content, and interactions. In theory, an effectively designed and implemented preventative solution should reduce the need for detection or incident response down the line. Interoperability is a key criterion for effective preventative solutions because they need to be able to work across a range of digital environments to secure content pipelines. Some key preventative measures that financial institutions can take to mitigate deepfake and disinformation risks are listed below.

- **Implement and enhance multi-factor authentication (MFA):** In the past, voice recognition was considered a highly secure authentication protocol, and some banks relied solely on it for authentication. However, with the advancement of AI voice generators and voice manipulation technologies, it is crucial to incorporate additional factors for customer, account, and employee verification. Options for enhanced authentication measures could include requiring employees to plug in a physical device or enhanced geolocation to flag suspicious transactions for additional verification.

- **Explore content provenance, transparency and authenticity solutions:** Another way that financial institutions can strengthen authentication protocols is by using secure content provenance technologies for any content-driven authentication process (e.g., upload a photo of a driver's license). The Coalition for Content Provenance and Authenticity (C2PA), founded by Microsoft, Adobe, Truepic, BBC, Arm and other industry members, has published an open standard on how to secure content using content provenance using cryptography.

- **Proactively protect high profile individuals within your organization**: As a supplement to technical authentication strategies, financial institutions can also proactively identify high profile individuals within their organizations who may be at higher risk of being imitated or impersonated via deepfake. One of our interviewees recommended that companies create and maintain secure databases of copyrighted images and video of their executives, captured using the C2PA's open standard. In the event of a deepfake incident, these repositories can then help to provide greater legal recourse.

- **Enhance standard employee training:** For banks, it's essential to provide internal training to enhance employee awareness and preparedness, given that deepfakes and generative AI are relatively new technologies. Making sure that employees understand how these technologies work and the risks associated with them is crucial to building preventative organizational resilience. Broader training can also be a great opportunity to better understand how deepfakes will impact employees at all levels of your organization. Operators in the call center, for example, could receive special training on deepfake audio to keep fraud risks top of mind and inform preventative protocols. These teams may also have unique insights as to how to better implement preventative mechanisms into the processes that they manage day to day.

- **Coordinate closely with PR, branding, marketing partners:** Because disinformation and deepfake threats take us outside of the 'traditional' cybersecurity realm of information security it is imperative to have clear lines of communication and regular touch points with cross-functional partners in PR, branding, marketing, and other verticals. Deepfakes and disinformation often travel at speed across public media channels. Close coordination with cross-functional partners can help to strengthen upstream resilience.

- **Invest in media literacy efforts:** Promoting media literacy across society can raise awareness of information protection and indirectly support your customers in discerning real from fake, fact from fiction. The SIFT media resiliency framework[41] can be promoted, which encourages individuals to "Stop, Investigate the information's source, find trusted coverage, and trace the original content." Banks can also educate their customers and clients on how data can be manipulated and encourage them to be cautious.

- **Engage generative AI providers in industry best practices and encourage corporate responsibility.** Commercial generative AI providers also have a crucial responsibility when it comes to preventing the misuse of their technologies for disinformation, fraud, influence operations, brand degradation, impersonation, and a host of other possible abuses. As an example, Synthesia has recently hired a specialized four-person team that focuses on preventing the use of its deepfakes for illicit or adversarial use and the spread of misinformation, particularly those containing hate speech or slurs.[42]

Prevention Solutions Providers: The C2PA & Its Members

Several of the C2PA's member organizations like Microsoft, Adobe, and Truepic provide solutions to help capture, secure, and preserve content provenance data (where, when, how a piece of content was created). For example, Truepic provides secure camera technologies to insurance companies, lenders, warranty providers and others to secure and digitize claims processes. Users can take a picture or video within a secure platform so that metadata is transparent and tamper-evident.

Truepic[43] and Adobe[44] also both provide signing solutions that allow content creators and owners to digitally 'sign' content using the C2PA's open standard and have both recently released examples of signed, AI-generated content that discloses the fact that it is synthetic in keeping with the Partnership on AI's framework on synthetic media. This kind of signing can also help brands secure their consumer-facing digital content (e.g. official communications, media, etc.). In March

[41] "Research Guides: Fake News and Information Literacy: The SIFT Method." The SIFT Method - Fake News and Information Literacy - Research Guides at University of Oregon Libraries,
[42] Satariano, Adam, and Paul Mozur. "The People Onscreen Are Fake. the Disinformation Is Real." The New York Times, The New York Times, 7 Feb. 2023,
[43] "Revel." Truepic, 5 Apr. 2023
[44] Content Authenticity Initiative (CAI). "Learn about Content Credentials," November 16, 2022. https://helpx.adobe.com/content/help/en/photoshop/using/content-credentials.html.

2023, the Center for Strategic and International Studies (CSIS) announced that it will be adopting the C2PA's open standard for its official image and video content.[45] C2PA powered cryptographic signing makes it easier to share with viewers official ownership of images and videos. It also helps to verify a piece of content's digital 'chain of custody,' which can support digital forensics down the line.

Obstacles for Prevention Strategies

Because the digital economy is increasingly interconnected, cooperation within and between industries is crucial, when it comes to prevention strategies. Open standards for content authenticity like the C2PA, can make this kind of collaboration easier, but ultimately require wider adoption by more companies and organizations. New technology adoption, even of an open standard, can take time, especially when decision-makers have limited technical vocabulary and knowledge. Media literacy education also takes time and can be a generational effort. There is also a risk that new authentication processes can introduce more friction into business operations or customer journeys.

Detection

There are a variety of providers for different forms of content. Throughout our research process, we were able to identify a few private sector organizations whose products and services aligned with financial institutions regulatory compliance obligations, as well as their protective needs and concerns. One example is Inscribe. Inscribe focuses on text/document detection and advertises a 50% reduction in application review time with over $80 million in fraud caught per month.

The detection of deepfakes is usually a combination of several techniques, which include deep learning, image and video analysis, audio analysis, and human expertise. At the moment, the leading technologies in this field include: GLTR, which is used to detect forged text using artificial intelligence, created and developed jointly by MIT-IBM Watson Laboratory and AI HarvardNLP. IBM has also developed a number of solutions, such as Watson Studio and Debator, to combat deepfakes and cyber threats.[46]

However, deepfake detection remains difficult. To begin with, detection necessitates a large amount of data: Deepfakes detection, as a subfield of deep learning, typically necessitates a training data set with a sufficient amount of data to ensure reliability. Furthermore, as deepfakes technology evolves, so must corresponding detection technologies, which necessitates the use of more and more complex data to ensure that these detection tools can continue to perform detection tasks effectively. Second, current detection methods have not yet been automated, and detection reliability has not met expectations, making large-scale deepfakes detection difficult.

[45] We Hold These Truths: How Verified Content Defends Democracies. "We Hold These Truths: How Verified Content Defends Democracies," n.d. https://www.csis.org/analysis/we-hold-these-truths-how-verified-content-defends-democracies.
[46] MIT-IBM Watson AI Lab. "Home - MIT-IBM Watson AI Lab," n.d. https://mitibmwatsonailab.mit.edu/.

Blackbird.AI is a company that works towards empowering trust, safety, and integrity across the "global information ecosystem". [47] It has a specific focus on the financial sector where it focuses on detection, analysis, planning, and mitigation when it comes to disinformation campaigns. The company believes in preventative practices to protect brand image, while simultaneously recognizing the ever-existing possibility of false narratives becoming headlines on social platforms. One way in which it specifically focuses on detection is through its Constellation Dashboard. This platform automatically surfaces emerging narratives and later uses threat intelligence to score narratives relating to toxicity, polarization, or bot-like networked activity. [48]

Providers that focus on audio detection, which is extremely relevant to the financial sector, include but are not limited to Eduworks (REVA), Pindrop, and Reality Defender. Eduworks is a detection tool for audio tampering, manipulation or deepfakes. It provides real time alerts to support efficient tracking and response. The focus on audio deepfakes are particularly important for account protection within the financial sector. Pindrop is a leader in voice authentication to prevent deepfake audio or voice spoofing. Pindrop has an established track record within the industry, especially with banks as it offers live detection for anti-fraud.

Video content detectors focus on deepfakes. Vendors include but are not limited to Intel FakeCatched, FaceForensics++, and DuckDuckGoose. Intel FakeCatcher offers real-time detection and thus far has claimed a 96% accuracy rate.[49] FaceForensic++ is an open access deepfake dataset. It is backed by Google and Jigsaw with a focus on helping people develop detection technology while training detection models effectively. [50] DuckDuckGoose is a start-up that uses machine learning to identify manipulated images or videos and claims an accuracy rate of 95%. The technology focuses on 3 types of deepfakes, FaceSwap, StyleGANs and Deep Puppetry. It has been used by the Dutch House of Representatives and the Netherlands Forensic Institute. [51]

Although the above-mentioned vendors have focused on specific content types, this does not mean that there are not some vendors that focus on multiple content types. Sensity AI is considered to be the world's first deepfake detection platform. It offers a suite of services such as live face detection, face matching, ID documents, and deepfake detection technologies to authenticate visual or text deepfakes. [52]

Reality Defender is another solutions provider that uses AI to provide detection of deepfakes before a problem arises. Our team was able to speak with the team Reality Defender to gather insight. Benjamin Colman (Co-Founder) and Matthew Banks (Strategy and Business Development Team) were able to inform us that the Reality Defender platform is able to

[47] Blackbird.AI: Mission. "Blackbird.AI: Mission," n.d. https://www.blackbird.ai/company/mission.
[48] Blackbird.AI: Solutions. "Blackbird.AI: Solutions," n.d. https://www.blackbird.ai/solutions.
[49] Demir, Ilke. "Intel Introduces Real-Time Deepfake Detector." Intel, 15 Nov. 2022,
[50] "Papers with Code - Faceforensics++ Dataset." Dataset | Papers With Code, 2022,
[51] "Detect Deepfakes Using Our Software." DuckDuckGoose
[52] "Biometrics KYC Verification Online - Sensity AI." Sensity, 20 Feb. 2023

specifically target deepfakes on the Internet by tracking narratives, analyzing propagation and deciding on how to strategically respond based on the actor.

Response

When it comes to addressing the issue of the response, the path is much more diffuse than what has been seen so far. The immense range of possibilities with which the available technology can be weaponized against companies in general -and the financial sector in particular- makes it almost impossible to assemble and implement a universal response protocol. In some cases, responding to the situation feeds the problem since it turns the attacker into a valid interlocutor; in other cases a quick response can cut the spiral that may have been triggered.

The response, as we concluded after having done the necessary research and having talked with the sources that we included in this report, must be designed and adapted according to the context in which the risk situation occurs and there are certain criteria that must be taken into account as the situation evolves. The creation of baseline response plans is possible, but the evolving threat landscape suggests that cyclical reviews and changes will be necessary.

According to Matthew F. Ferraro, a victim company may seek to cooperate with social media platforms to request aid in preventing the spread of false narratives. A company should also respond to false speech with truthful, positive messaging about itself and consider publicly announcing that it is being targeted by disinformation. After that, companies should consider contacting their regulators, shareholders, customers, and partners, they should also consider pursuing legal action against deceivers if required and appropriate.[53]

Hereinafter cases that are considered essential for a response are presented. However, it is noted that these guidelines are not a recipe or an organic way to respond. They are only a compendium of notes and information that should be taken into account based on the experience, research and opinions that we have gathered for the preparation of this document:

VI) Case Study: Silicon Valley Bank

In a report[54] published in 2021 by Price Waterhouse Cooper (PWC) it is indicated that, until now, people with political or social influence have been the main victims of malicious information campaigns; however, financial institutions and corporations have begun to be targeted by bad actors to harm their digital assets and reputation.

[53] "Disinformation and Deepfakes Risk Management (DDRM) - Fake Viral Narratives and Synthetic Media Pose Risks to Business | JD Supra."
[54] Upton, Philip, et al. "Disinformation Attacks Have Arrived in the Corporate Sector. Are You Ready?" PwC, 9 Feb. 2021

Our interview with disinformation mitigation service provider, Alethea, allowed us to further explore the implications that misinformation and disinformation had on the Silicon Valley Bank's collapse. As disinformation thrives under crisis, Alethea was able to successfully track how social media platforms furthering misinformation and disinformation campaigns were able to encourage online panic and later leading to clients removing their money from SVB leading to the bank's failure. [55]

Alethea found that not only were venture capitalists fueling panic, but foreign state media such as Chinese and Russian state propaganda outlets played a part in spreading the false information. [56] Additionally, foreign propaganda sites directly linked to disinformation campaigns but a far-right conspiratorial financial blog, Zero Hedge had published over 20 articles on March 13th about SVB's collapse on platforms such as Reddit, Truth Social, and Twitter. [57] Although SVB's collapse was not directly caused by disinformation campaigns, the spread of this information instilled fear in SVB's client base.

In the case of SVB, it was largely reactive. The propaganda being shared was also domestic and was largely right-wing alarmist sites claiming that the bank had failed due to leftist sentiments, said one of the team members at Alethea. With the most impactful events coming from private chats, the visibility into the campaigns was low, leading to further confusion. The role of generative AI and emerging tech was indisputable in this case. The information may not have been altered in some cases, lots of influence operations often contain facts or partial truths, but the speed at which it spread was a novelty compared to the 2008 collapses.

When the bank run began to affect SVB, the amount of information directed at its clients multiplied rapidly. In this regard, Bloomberg indicates the following: "Silicon Valley Bank lost $42 billion (…) within hours when its run began. Regulators and executives have noted social media's role. Citigroup Inc. Chief Executive Officer Jane Fraser recently called the combination of mobile-banking apps and social media a game-changer for bank stability when discussing SVB."

Another thing Alethea noticed was the prevalence of scammers, particularly crypto scammers, that utilized the public panic emerging from these fear mongering narratives to influence people in turning to their crypto-alternatives. Since people were starting to panic, these scams worked well in the context of a bank run and were readily accepted.

In a March 2023 note, Bloomberg[58] reviews a number of cases—ranging from the political to the financial—in which deepfakes have been weaponized. In both the case of Silicon Valley Bank (SVB) and Credit Suisse that the Bloomberg report addresses, there is no direct use of synthetic media. It is real information that was disseminated -perhaps pursuing a pernicious purpose- through social networks in an accelerated manner.

[55] "When Crisis Strikes, Disinformation Thrives." Alethea, Alethea, 15 Mar. 2023,
[56] Ibid
[57] Ibid
[58] Davies, Paul J., and Parmy Olson. "Online Fakery and Digital Bank Runs Are a Scary Mix." Bloomberg.com, Bloomberg, 28 Mar. 2023

The Credit Suisse case is one where social networks also played a leading role. Credit Suisse suffered not a financial crisis, but years of accumulating reputational scandals that were amplified by the media and shared on social media. Thus, when a tweet indicated that a European bank was on the verge of collapse, users did not hesitate to withdraw their money immediately and together.

VII) Conclusion

In conclusion, the growing sophistication and accessibility of deepfakes pose a significant threat to the financial sector, as cyber threat actors can use them to impersonate key personnel, manipulate markets, and conduct fraudulent activities. Given the current geopolitical climate, the risk of disinformation campaigns has also increased, fueled by generative AI. Information operations are an effective tool to influence population behavior, which is relevant to the health of financial institutions. Market predictability, or monitoring, which are vital to the health of financial institutions, has become increasingly difficult. The findings and insights from our research illuminates the need for financial institutions to adapt and boost their security strategies and implement measures to prevent, detect, and respond to deepfake and disinformation threats.

Employee training is a critical pillar for enhancing organizations' resilience and preparedness to these threats, especially for high-profile executives susceptible to being impersonated through deepfakes or spear phishing campaigns. Personal training for those dealing with transactions regularly operators can also play a significant role in safeguarding valuable customer data, particularly in the event that detection technology were to be introduced in the institution’s security ecosystem. Media literacy training and workshops could also keep staff at all levels aware of the current information landscape and to continuously question whatever they see or hear.

Multi-factor authentication is also an essential practice to ensure authenticity, given that deepfakes are often undetectable to the human eye. While potentially user unfriendly to certain staff members or senior executives, multi-factor authentication provides an additional layer of security and identity verification needed today. The report provides recommendations for mitigation strategies, including prevention, detection, and incident response, to improve the financial sector's defenses against deepfakes and other disinformation threats.

Continued evaluation of these technologies and the risks they pose will be essential, not just for the financial sector, but private industry at large. As regulatory efforts slowly emerge to attempt to impose guidelines that match these threats, it is imperative for financial institutions and other private sector players to take the lead on potential strategic responses to mitigate the risks posed by these new technologies.

The recommendations provided in the report offer practical solutions for financial institutions to enhance their cyber defenses and protect against deepfake and disinformation threats. By

adopting these measures, financial institutions can safeguard their brand reputation, customer trust, and operations, and remain resilient against emerging cyber threats.

Acknowledgements

We would like to express our gratitude to Neal Pollard and Beth Cartier for their invaluable contributions as capstone advisors, providing us with insightful feedback and guidance. Without their exceptional leadership and unwavering support, our project would not have achieved the same level of progress and success.

References

Ajder, Henry. The Ethics of Deepfakes Aren't Always Black and White. 16 June 2019, https://www.proquest.com/docview/2253189767?pq-origsite=primo.

Ajder, Henry, et al. "The State of Deepfakes." DarkTrace, https://regmedia.co.uk/2019/10/08/deepfake_report.pdf.

"The American Public Views the Spread of Misinformation as a Major Problem - AP-NORC." AP NORC University of Chicago, 8 Oct. 2021, https://apnorc.org/projects/the-american-public-views-the-spread-of-misinformation-as-a-major-problem/.

Allyn, Bobby. "Deepfake Video of Zelenskyy Could Be 'Tip of the Iceberg' in Info War, Experts Warn." NPR, NPR, 17 Mar. 2022, https://www.npr.org/2022/03/16/1087062648/deepfake-video-zelenskyy-experts-war-manipulation-ukraine-russia.

Bateman, Jon. "Deepfakes and Synthetic Media in the Financial System: Assessing Threat Scenarios." Carnegie Endowment for International Peace, July 2020, https://carnegieendowment.org/files/Bateman_FinCyber_Deepfakes_final.pdf.

"Biometrics KYC Verification Online - Sensity AI." Sensity, 20 Feb. 2023, https://sensity.ai/.

Bishop, Adrian. "How Misinformation Can Impact Businesses." TechInformed, 6 Sept. 2022, https://techinformed.com/how-misinformation-can-impact-businesses/.

Blackbird.AI: Mission. "Blackbird.AI: Mission," n.d. https://www.blackbird.ai/company/mission.

Blackbird.AI: Solutions. "Blackbird.AI: Solutions," n.d. https://www.blackbird.ai/solutions.

"Build Software Better, Together: Voice Cloning." GitHub, 2023, https://github.com/topics/voice-cloning.

Chesney, Robert, and Danielle Keats. "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security ." 107 California Law Review 1753 (2019), U of Texas Law, Public Law Research Paper No. 692, U of Maryland Legal Studies Research Paper No. 2018-21, Http://Dx.doi.org/10.2139/Ssrn.3213954, 14 July 2018.

Content Authenticity Initiative (CAI). "Learn about Content Credentials," November 16, 2022. https://helpx.adobe.com/content/help/en/photoshop/using/content-credentials.html.