What executives need to know and do about deepfakes

Threats to Contact Centers and the Solution

> The difference between false memories and true ones is the same as for jewels: it is always the false ones that look the most real, the most brilliant.
>
> **– Salvador Dali**

Section 1

Introduction - The rise of deepfakes

Deepfakes have recently attracted significant attention in the news, especially since the emergence of cutting-edge generative artificial intelligence tools. Company boards and executives are beginning to recognize how these impact their business and identify essential initiatives required for safeguarding their operations and customers.

Contact center fraud grew by 40% in 2022 Year-over-Year and this trend is expected to continue in 2023[1]. Since 2020, data breaches have affected over 300 million victims causing over $8.8 billion in reported losses[2], and have led to negative impact on brands and customer trust. Deepfakes and synthetic voice attacks are a key element of this growing fraud problem causing concerns for company leadership and consumers alike.

Contact center technology teams are actively assessing the threat posed by deepfakes to effectively address these concerns. Pindrop can help contact center teams to stop this threat. Pindrop has proven technology with demonstrated excellence in stopping fraud with multifactor, real-time, cloud native fraud detection across all stages of the call center. Forrester Consulting's "Total Economic Impact (TEI)" study[3] found that Pindrop customers experienced an average of 171% Return on Investment (ROI) including over $6 million in fraud reduction savings over a three year period, with the ability to detect 15% more fraud in addition to other systems.

Read on to learn how you can proactively prevent, detect and eliminate the threat of deepfakes and protect your business and your customer.

What are deepfakes?

To start, generative AI is a form of artificial intelligence that produces synthetic digital content, typically text, images, videos or audio. The term deepfake describes digitally manipulated content that convincingly replaces one person's voice, image, and/or likeness with that of another.

Deepfakes, particularly audio-based ones, are everywhere in the news:

- Senator Blumenthal's opening remarks at the Senate hearing on AI[4]
- Wall street journal reporter Joanna Stern's AI challenge[5]
- Donotpay's CEO's call to ask for a refund from a large bank[6]
- How I broke into a bank account with an AI-generated Voice[7]

Section 2
How are deepfakes a threat to my business

In the current business landscape, organizations are increasingly relying on remote customer interactions to deliver higher value services, such as executing stock market trades, registering for telemedicine, accessing their bank accounts, and more. In these interactions, businesses face the challenge of relying on remote identity confirmation. Audio deepfakes present yet another challenge during this confirmation process. Now, bad actors can use deepfakes to run targeted campaigns impersonating real customers. Let's explore where your contact center is vulnerable.

[4] https://www.cnn.com/videos/business/2023/05/16/artificial-intelligence-hearingblumenthal-ai-voice-nc-vpx.cnn
[5] https://www.wsj.com/video/series/joanna-stern-personal-technology/24-hour-challengecan-my-ai-voice-and-video-clone-replace-me/EC817295-03D0-4031-B40B-694D7BDE2797
[6] https://www.vice.com/en/article/pkg94v/deepfake-voice-do-not-pay-wells-fargo-refund
[7] https://www.vice.com/en/article/dy7axa/how-i-broke-into-a-bank-account-with-an-aigenerated-voice

4 known deepfake attack methods:

Recorded voice fraud involves attackers using obtained voice recordings (with entirety of speech or with concatenated parts) to deceive voice biometric authentication systems, gaining unauthorized access to accounts in call centers.

Speech synthesis refers to the creation of a voice model by fraudsters who generate spoken words from text that convincingly resemble those of a real person. Examples of AI engines used for speech synthesis are ElevenLabs, ResembleAI, and Vall-E.

Voice chatbot fraud involves fraudsters utilizing an automated chatbot and voice model to mimic the sound and behavior of another real person during interactions. Examples of AI engines that offer voice chatbot are Google Duplex and DoNotPay.

Voice conversion is where an attacker alters their voice in real-time to resemble that of an actual person by speaking into a model that takes voice as an input. An example of an AI engine that offers such a voice conversion is Respeecher.

Five ways deepfake attacks can hurt your business:

1. Increase in fraud losses
2. Damage to reputation and brand
3. Breach of sensitive customer data
4. Increased operational costs and disruptions
5. Loss of customer trust and potential customer churn

Additionally, although compliance requirements specific to deepfakes are still a work-in-process, it's not far-fetched to expect incremental compliance requirements for regulated industries, especially those that already have requirements for customer authentication.

The congressional inquiries to top US banks[8] illustrate the concerns among regulators and policymakers.

[8] https://www.banking.senate.gov/newsroom/majority/brown-presses-banks-voice-authentication-services

Section 3

How are deepfakes made?

"Synthetic media has become far easier to create with the roll out of the so called generative AI systems that can quickly transform simple videos into sophisticatedlooking videos, photos, music and text", as noted by WSJ in a recent article[9]. Bad actors can create a deep fake at almost no cost, using the freemium service offered by text-to-speech (TTS) services such as ElevenLabs and ResembleAI.

To get started, a bad actor will collect the victim's voice recording either from publicly available videos or audio recordings, such as YouTube or a podcast, or from recording a phone conversation which can be easily done, for example, in the pretext of a marketing survey. TTS services such as Microsoft's Vall-e claim to create a deepfake with as little as 3 seconds of a person's audio[10]. The bad actor uploads the victim's voice to the TTS service to create a custom deepfake model of the victim's voice. Finally, the bad actor prepares a set of phrases that can be used during a phone conversation (e.g. "My voice is my password")[11]. The TTS service uses the phrases to produce the corresponding audio recordings.

In recent test results, Pindrop classified deepfakes generated from Microsoft's Vall-e Text-to-Speech system with a 99% accuracy.

Can deepfakes be detected? If so, how?

There are multiple efforts in the technology industry and the AI community to develop an authentication infrastructure for digital content, e.g., tools that will mark AI-generated content with data about its origin. Unfortunately, these efforts are not universally adopted, especially given the proliferation of opensource AI models. In other words, the deepfake genie is already out of the bottle. The proliferation of generative AI models has already given bad actors access to tools to create deepfakes. So, the onus of detecting deepfakes is now on individual businesses. The good news is that deepfakes can be detected with a high degree of confidence, especially when businesses use deepfake detection as part of a multi-factor authentication platform, such as Pindrop.

[9] https://www.wsj.com/articles/ais-rapid-growth-threatens-to-flood-2024-campaigns-with-fake-videos-dbd8144fmod=Searchresults_pos1&page=1

[10] https://arstechnica.com/information-technology/2023/01/microsofts-new-ai-can-simulate-anyones-voice-with-3-seconds-ofaudio/

[11] https://www.vice.com/en/article/dy7axa/how-i-broke-into-a-bank-account-with-an-ai-generated-voice

Deepfake detection works on the basic premise that any deepfake generator creates artifacts and patterns in the various acoustic elements of the generated audio such as frequency, voice variance, pauses, etc, which are distinctly different from natural human speech. These patterns may not be detectable by human ear, but can be identified when analyzed by specially designed artificial intelligence tools.

Pindrop's deepfake detection technology relies on analyzing these artifacts and patterns. Through the processing of these artifacts and patterns, Pindrop extracts proprietary spectro-temporal features from the audio, what Pindrop calls 'liveness' objects, and then analyzes them using a deep neural network (DNN) model. Pindrop's DNN model is a proprietary model built over 8 years, designed specifically for recognizing genuine human voice. The unique architecture of this robust model empirically shows unmatched levels of accuracy and generalization[12]. Pindrop's liveness objects are then used to estimate a 'liveness' score. This capability allows Pindrop to protect zero-day attacks from deepfakes generated by new models, not previously seen.

As an example, Pindrop detected several deepfakes in the public domain, such as Anthony Bourdain's deepfake voice in Roadrunner in 2021 (article)[13], Senator Blumenthal's opening remarks at the Senate hearing on AI in 2023 (blog post)[14]. In recent test results, Pindrop classified deepfakes generated from Microsoft's Vall-e Text-to-Speech system with a 99% accuracy[15]. Further, Pindrop's deepfake detection technology has been tested and benchmarked by both academic and institutional groups on various types of deepfake attacks[16].

Section 4

Can voice biometrics outsmart deepfake attacks?

Generative AI is improving at a rapid pace, raising concerns if detection tools can stay ahead and protect customers especially against zero-day attacks. Recent press articles like the one in New York Times[17], have created the impression that Generative AI deepfakes can readily pass biometrically based authentication methods. This is not the case.

[12] Generalization Of Audio Deepfake Detection, Tianxiang Chen, Avrosh Kumar, Parav Nagarsheth, Ganesh Sivaraman, Elie Khoury, November 2020

[13] https://www.wired.com/story/these-hidden-deepfakes-anthony-bourdain-movie/

[14] https://www.pindrop.com/blog/pindrop-deepfake-detection

[15] This was conducted on an internally curated dataset of about 200k samples of both genuine and deepfake samples

[16] Pindrop research team's submission to the ASVspoof 2021 challenge, Tianxiang Chen, Elie Khoury, Kedar Phatak, Ganesh Sivaraman, 2021

[17] https://www.nytimes.com/2023/05/18/technology/ai-chat-gpt-detection-tools.html

Even if deepfake attacks were to increase, voice biometrics will continue to be a viable and preferred authentication technology for enterprises, for the following reasons:

1. Each person has distinct voice characteristics, including pitch, tone, cadence, and pronunciation. Voice biometrics analyze these unique patterns, making it difficult for voice cloning to perfectly replicate all aspects of an individual's voice.

2. Using voice biometrics in conjunction with other authentication factors, such as device and behavior authentication, creates a multi-factor system that is harder for attackers to bypass. Organizations must take the defense-in-depth approach to ensure multiple factors are validated before access is granted to the user.

3. It is not only possible to detect synthetically generated voices using advanced AI and machine learning (ML) techniques such as deep neural networks, but it can also be done in real-time. Pindrop has been working on this technology for over 8 years and has multiple patented technologies and techniques to detect even the most sophisticated deepfakes in real-time. in summary, the difference is the biometric solution leveraged.

Overall, the emerging threat of deepfakes emphasizes the importance of continually advancing and strengthening voice biometric systems to mitigate these evolving risks and ensure their reliability in providing robust authentication and fraud detection.

How can you stay ahead of deepfake attacks?

As with any good security program, businesses need to take a holistic, multi-pronged approach to protecting against deepfakes which should cover both human elements and technology capabilities. The good news is that most companies have well established security and fraud programs within the contact center as part of the overall security program for the business. These programs can be extended to protect against deepfakes.

Management teams view deepfake as a genuine threat, but one that can be effectively managed. To stay apprised of deepfake threats, boards should ensure the following are in place:

1. A dedicated initiative to build deepfake protection capabilities in the company, with the necessary budget and resources.

2. Deepfake detection capabilities embedded into the voice authentication and anti-fraud technology platforms and contact center processes.

3. A clear communication plan for educating customers on how they are being protected against deepfakes. Unlike other security vulnerabilities, deepfakes can affect a customer's experience with the brand.

4. Incorporation of deepfake detection into the Board's overall security review and oversight process.

What is Pindrop doing to protect its customers against deepfakes?

Deepfake detection is a core strength and differentiator for Pindrop. Pindrop's research team has invested 13 years into analyzing audio and voice signals to look for anomalies, find fraudulent activities and authenticate callers. Pindrop provides the most comprehensive voice security platform for detecting deepfakes across all channels. With focus on deepfake detection for 8 of the last 13 years, Pindrop has developed several patented technologies to detect even the most sophisticated of the synthetic voices.

Pindrop is set to launch advanced deepfake detection capabilities, which are meticulously trained on our extensive proprietary and public datasets[18]. These cutting-edge capabilities are expected to be available in the second half of 2023 and will be integrated into both Passport and Protect cloud platforms for all of our customers to leverage.

With focus on deepfake detection for 8 of the last 13 years,

Pindrop has developed several patented technologies to detect even the most sophisticated of the synthetic voices.

Section 5
How Pindrop customers benefit

Real-time deepfake detection

Similar to voice match capability, Pindrop's deepfake detection technology enables customers to get real-time feedback. This will allow enterprises to authenticate or reject the caller during the call. Pindrop can also advise on how companies can operationalize that process.

Fully integrated, cloud-native capability

The deepfake detection technology will be offered to existing customers of Pindrop Passport and Protect. New Pindrop customers will have the ability to leverage deepfake detection technology as part of their cloud platform.

Most comprehensive coverage

Pindrop's deepfake detection technology provides protection for not only the synthetically generated speech (text-to-speech, voicebot, voice modulation, voice conversion), but also the recorded voice playback attacks.

[18] As of June 2023, this dataset has about 6 millions samples covering various deepfake types

Industry leading performances

Pindrop's deepfake detection technology protects against the presentation attacks that use recorded voices, text-to-speech, cloned voice or voice conversion techniques. The detection rate for all presentation attacks is >92% with a False Rejection Rate (FRR) of genuine audio of 1% or less with netspeech of 2 seconds on an internally curated benchmark19. Additionally, on the ASVspoof 2019 dataset, the equal error rate is as low as 1.26%20.

Continuous enhancements for deepfake evolution

Similar to our fraud detection capabilities, Pindrop's deepfake detection technology is designed to allow continuous delivery and release of improvements to our customers to allow our customers to stay current on the latest release without having to worry about upgrades.

Explore Pindrop's deepfake detection technology at work in this video where Pindrop detects the deepfake used by Senator Blumenthal in his address to the Senate[21].

To schedule a demo for your team, please contact your Pindrop Account Executive or email us at deepfake@pindrop.com.

[21] https://www.pindrop.com/blog/pindrop-deepfake-detection

Pindrop helps contact centers detect fraud attempts throughout their organization by analyzing risk on live calls and customer accounts providing added protection against fraud attempts.

Pindrop solutions leverage real-time risk analysis for all inbound calls and monitor the contact center as well as the IVR and can help alert on at-risk customer accounts that show subtle signals of account takeover. Its precise voice identification technology recognizes unique identifiers within the human voice that can enable its customers to prevent more fraud and deliver exceptional customer experiences in call centers, obtain information from smart devices and even activate cars. A privately held company, Pindrop is venture-backed by Andreessen Horowitz, Citi Ventures, Felicis Ventures, CapitalG, GV, IVP, and Vitruvian Partners. Since its inception, Pindrop has analyzed more than 5 billion voice utterances, detected over 3 million fraud calls, and saved our customers more than 2 billion dollars and counting.

Visit pindrop.com for more information.

Pindrop®

Schedule a call with one of our experts today

pindrop.com | info@pindrop.com