# Liveness Detection for Audio Deepfake and Voice Clone Detection

## The Problem

Deepfakes have spread to the public domain with misinformation related to war and elections, as well as for committing fraud. Voice clones, a type of deepfake where a fraudster uses a voice and text model to generate spoken words that sound like an actual person, have already been used to target families, takeover bank accounts and to impersonate executives causing financial losses in millions of dollars.

With the advancement in generative artificial intelligence (GenAI), creating voice clones has become quicker and cheaper. Scammers can now utilize chatbots and scale up their attacks instead of employing several people for the same amount of effort. With the latest technology it takes only three seconds of an individual's voice, which is easily available on social media, to create a deepfake of that individual.

In addition to the risk of financial fraud and disinformation, voice clones also pose a challenge of authenticating genuine individuals and the level of trust that can be ascribed to each interaction.

## Our Submission

Our submission highlights the capabilities of our liveness detection technology to detect voice clones and audio deepfakes in real time, helping to prevent unauthorized users from accessing valuable assets and authenticate genuine users with a high level of trust. This method of liveness detection evaluates each incoming phone call or digital audio in real-time in 2-second chunks and computes a liveness score. If the score is lower than a predefined threshold, the audio is flagged as potentially deepfake (a.k.a. “cloned voice,” “spoofed voice”). Otherwise, the audio is considered likely “Live” (a.k.a. “real” or “human”). This method leverages recent breakthroughs in deep neural networks (DNN) and is trained internally on diverse training data involving over 120 text-to-speech and voice cloning systems. This comprehensive dataset makes our proposed solution resilient to unseen types of voice cloning techniques.

We have tracked the performance of our system over the past year on 17 different public and internal voice cloning benchmarks. While the baseline system that had the lowest Equal Error Rate (EER - where the false positive rate is equal to the false rejection rate) as a single system at the ASVspoof 2019 LA (Logical Access) task resulted in an average EER of 36.32% on those benchmarks, the newly proposed system achieves a low EER of only 4.02%. This solution is the culmination of 18 different patent applications (pending or granted) covering multiple aspects of voice liveness detection.

Our submission highlights the need for a comprehensive audio deepfake detection approach that covers all aspects of artificial voice, including voice recordings, synthetic speech, automated chatbots, and real time voice conversion while working in both 8 kHz and 16 kHz, channels with low latency in cloud, on-prem and on-device environments. The solution can be integrated with existing fraud detection and user authentication systems to facilitate easy deployment into existing infrastructure.