eBook

AI in the Shadows

Deepfake Risks, Detection, and Mitigation Strategies for Banking & Insurance

cpl.thalesgroup.com

THALES
Building a future we can all trust

Executive Summary:

Deepfakes have rapidly evolved from a novelty to a serious security concern, particularly for financial institutions. These AI-generated forgeries can create convincing fake videos, audio, and images that are increasingly difficult to distinguish from genuine content. This eBook examines the emerging risk of AI-generated deepfakes in the banking and insurance sectors. It explores how this technology is being used to perpetrate fraud, discusses methods for detecting deepfakes, and outlines strategies for mitigating associated risks.

Introduction

The Rise of Deepfake Technology

Deepfakes, hyper-realistic digital forgeries (or synthetic identities) created using artificial intelligence, have rapidly evolved from a novelty to a serious security threat. This technology can be used to create convincing fake videos, audio, and images that are increasingly difficult to distinguish from genuine content.

Implications for Banking and Insurance

The financial sector, particularly banking and insurance, is especially vulnerable to deepfake attacks due to its reliance on identity verification and the high stakes involved in financial transactions and claims processing.

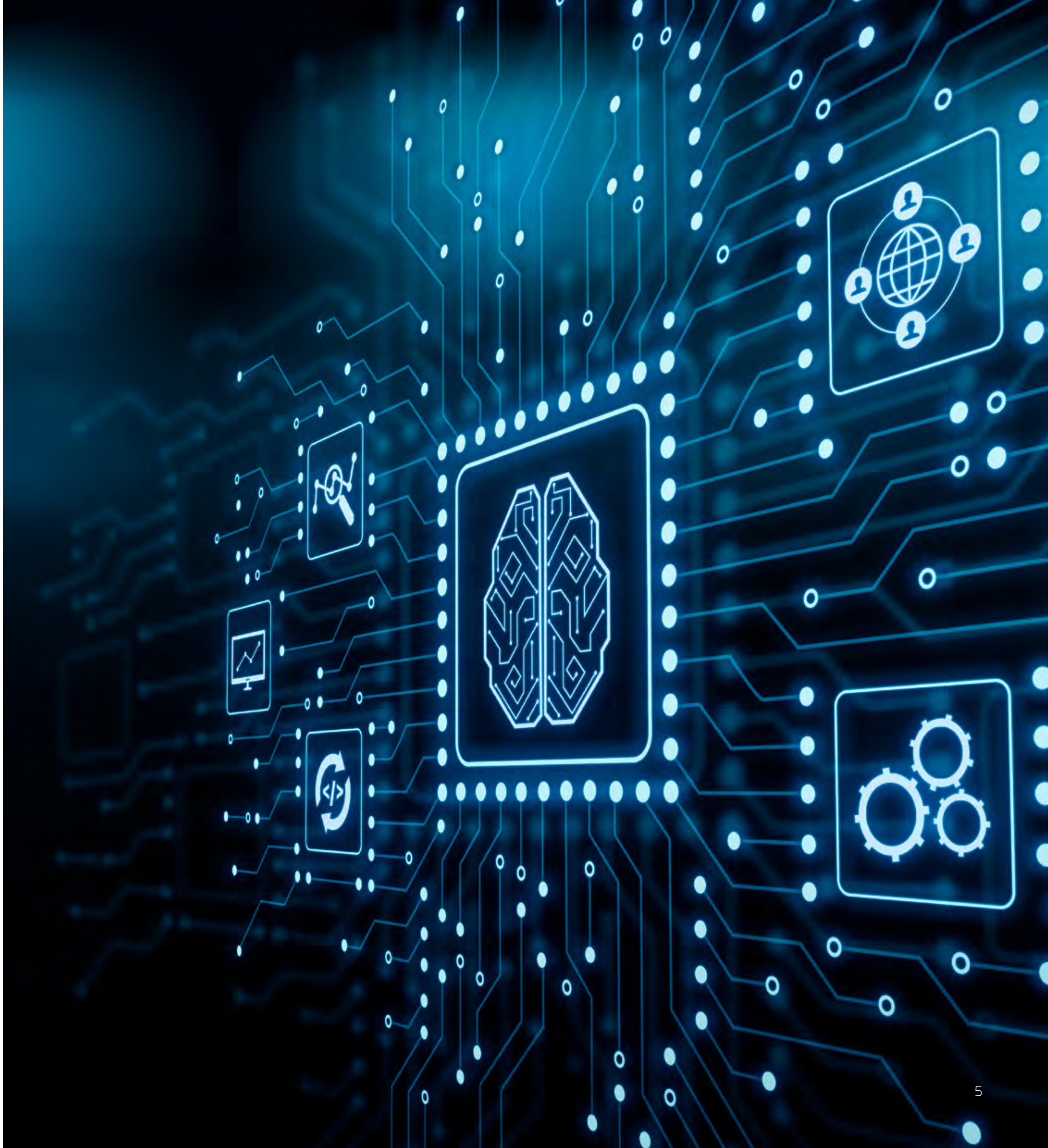
Understanding Deepfake Technology

How Deepfakes Work

Deepfakes are created using deep learning algorithms, particularly generative adversarial networks (GANs). These AI systems can analyze existing images or videos of a person to generate new, synthetic content that mimics their appearance and voice with startling accuracy.

Types of Deepfakes

- Video deepfakes: Synthetic videos where a person's face or entire body is replaced with someone else's
- Audio deepfakes: AI-generated voice cloning that can mimic a person's speech patterns and tone
- Image deepfakes: Artificially created or manipulated still images



Risks to Banking and Insurance

- **Identity Fraud**
  Deepfakes can be used to bypass biometric security measures, potentially allowing criminals to access accounts, apply for loans, or file false insurance claims under someone else's identity.

- **Social Engineering Attacks**
  Criminals can use deepfake technology to impersonate executives or employees, manipulating staff into transferring funds or revealing sensitive information.

- **Reputational Damage**
  Deepfakes could be used to create false statements or compromising situations involving company executives, potentially damaging the institution's reputation.

- **Insurance Claim Fraud**
  Deepfaked video or audio evidence could be used to support fraudulent insurance claims, making it harder for insurers to detect false claims.

Recent Examples of Deepfake Fraud

- **Case Study: UAE Bank Heist (2023)**
  In early 2023, fraudsters used deepfake technology to steal $35 million from a bank in the United Arab Emirates. The criminals used AI to clone the voice of a company director, convincing a bank manager to transfer funds as part of an acquisition.

- **Case Study: Crypto.com Deepfake Scam (2023)**
  Scammers created deepfake videos of Crypto.com's CEO to promote a fake "giveaway" scheme, leading to financial losses for victims who believed the fraudulent promotion.

- **Case Study: Binance KYC Deepfake Attempt (2022)**
  Binance, a cryptocurrency exchange, reported an increase in sophisticated deepfake attacks attempting to bypass their Know Your Customer (KYC) verification process.

So how do we prevent such attacks?

The Thales OneWelcome Identity Platform offers a comprehensive set of solutions to combat the growing threat of deepfake attacks in the banking and insurance sectors. This advanced platform integrates multiple layers of security and authentication to protect against sophisticated AI-generated fraud attempts starting with detection methods, discussed below, and deploying robust mitigation strategies.

Deepfake Detection Methods

- **Injection Detection**
  Injection detection focuses on identifying artifacts or inconsistencies introduced during the deepfake creation process. This can include:
  - Analyzing video metadata for signs of manipulation
  - Detecting inconsistencies in lighting, shadows, or reflections
  - Identifying unnatural blinking patterns or facial movements

- **Presentation Attack Detection**
  Presentation attack detection aims to differentiate between a real person and a deepfake presented to a biometric system. Techniques include:
  - Liveness detection: Ensuring the subject is a living person, not a static image or video
  - Texture analysis: Examining skin texture and other fine details that may be lost in deepfakes
  - Behavioral biometrics: Analyzing unique patterns in a person's movements or speech

Mitigation Strategies for Financial Institutions

Multi-Factor Authentication

Implement robust multi-factor authentication systems that combine something the user knows (password), has (device), and is (biometrics).

Liveness Detection and Behavioral Biometrics

Utilize cutting-edge biometric technologies that are more resistant to deepfake attacks, such as:

- 3D face mapping
- Behavioral biometrics (e.g., typing patterns, mouse movements)

By monitoring user behavior patterns such as typing speed, mouse movements, and navigation habits, Thales' solutions can identify anomalies that may indicate fraudulent activity.

Risk Management

Thales offers comprehensive risk management tools to continuously monitor and mitigate potential threats:

- Advanced Fraud & Risk Management: Thales employs machine learning algorithms to detect unusual patterns and flag suspicious activities in real-time, helping institutions stay ahead of evolving fraud techniques.
- Compliance and Security Protocols: Thales ensures that their solutions comply with regulatory requirements and are regularly updated to address new security challenges.

Employee Training

Educate staff about deepfake threats and train them to recognize potential signs of manipulation in video or audio communications.

Invest in Deepfake Detection Technology

Deploy AI-powered deepfake detection tools as part of the security infrastructure, particularly for customer onboarding and high-value transactions.

Collaboration and Information Sharing

Participate in industry-wide initiatives to share information about deepfake threats and best practices for detection and prevention.

Conclusion

As deepfake technology continues to advance, the risks to the banking and insurance sectors will likely increase. Financial institutions must stay ahead of this threat by investing in detection technologies, implementing robust security measures, and educating their staff and customers. By taking a proactive approach to deepfake mitigation, banks and insurers can protect themselves and their clients from this emerging form of fraud.

The Thales OneWelcome Identity Platform offers a comprehensive and adaptable solution to the growing threat of deepfake attacks in the financial sector. By combining advanced biometrics, AI-powered detection, risk-based authentication, and continuous monitoring, the platform provides a robust defense against even the most sophisticated deepfake attempts. As deepfake technology continues to evolve, Thales's commitment to innovation and security ensures that financial institutions can stay one step ahead of potential fraudsters, protecting both their assets and their customers' trust.

THALES

Building a future we can all trust

Contact us

For all office locations and contact information,
please visit cpl.thalesgroup.com/contact-us

cpl.thalesgroup.com