Corporate Illusions: Unmasking the risk of deepfakes in the boardroom and beyond.

Tallulah Midnight Marin O'Hanlon

A dissertation submitted in part-fulfilment of the degree of

BA (Hons) Criminology

Nottingham Trent University

March 2025

Abstract

Deepfake technology, powered by artificial intelligence and deep learning, presents a growing threat to corporate environments. This study fills a crucial research gap examining the risks deepfakes pose to companies, investigating (1) employee and executive awareness of deepfake threats, (2) human ability to distinguish between real images and deepfake images, (3) the primary risks that deepfakes pose to companies, and (4) corporate preparedness and response strategies. A mixed-methods approach was employed, combining a quantitative survey of 229 corporate employees with qualitative insights from nine cybersecurity professionals. The survey assessed employee's ability to distinguish between real and deepfake images, while interviews provided expert perspectives on deepfake threats, corporate vulnerabilities and mitigation strategies. Thematic analysis found five key themes, (1) Public Perceptions, (2) Current and Future Threats of Deepfakes, (3) Corporate Preparedness and Response Measures, (4) Risks Associated with Deepfakes for Companies, and (4) Measures companies have put in place. While executives exhibit greater awareness of deepfake risks, this does not affect accuracy rates. Corporate preparedness remains low, and many companies still rely on traditional cybersecurity frameworks, lacking specialised deepfake detection tools or employee training. The research also highlights the rapid advancement of deepfake technology, the growing accessibility of deepfake tools, and the increasing financial and reputational risks associated. Furthermore, statistical analysis found that employees significantly struggle to detect deepfakes, with an average accuracy of 32.36%, far below the 53.16% benchmark from previous studies. The study concludes that deepfakes represent an urgent and under-recognised corporate security threat. To mitigate these risks, the study proposes the **D.E.T.E.C.T** framework to help companies be proactive rather than reactive to deepfake threats: **D**eploy detection tools, **E**valuate weaknesses and vulnerabilities across the company, **T**rain employees and strengthen company trust, **E**nforce guidelines ensuring crisis response strategies are strong, **C**ommunicate ways for employees to report suspicious media, and **T**rack the deepfake threat landscape to stay on top of defences.

Table of Contents

List of Tables

List of Figures

Glossary

As defined from the Oxford English Dictionary (2025).

Word	Definition
Artificial Intelligence	The capacity of computers or other machines to exhibit intelligent behaviour. Software that performs tasks previously thought to require human intelligence.
Corporate Company	Large company or business corporation
Deep Learning	A type of machine learning considered more dynamic
Deepfakes	Various media sources that have been digitally manipulated to replace a person's likeness convincingly with that of another.
Machine Learning	The capacity of computers to learn and adapt without following explicit instructions by using algorithms and models to analyse and infer from patterns in data
Misinformation	Wrong or misleading information
Social Engineering	The use of deception to induce a person to divulge private information or unwittingly provide unauthorised access to a computer system or network

1. Introduction

Imagine joining a video meeting where your CEO announces mass layoffs, admits to financial fraud, or declares bankruptcy. Instant panic spreads through the office, messages pinging with fear, employees scrambling for answers, and the stock plummeting to nothing. Only to realise that it was entirely fabricated by artificial intelligence (AI). This is not a dystopian future, it is the reality of deepfake technology (ADL, 2023; Agarwal, et al., 2019; Bateman, 2022; Flick & Morehouse, 2011; Miller, 2021; Munk, 2024; Salahdine & Kaabouch, 2019; SumSub, 2024). Deepfake technology, powered by AI and deep learning, has advanced to a point where it is impossible to tell the difference between real and synthetic media (Karras, et al., 2020; Karras, et al., 2024; Karras, et al., 2020). While initially a tool for entertainment (Bregler, et al., 2023; Masood, et al., 2023; Tolosana, et al., 2020), deepfakes now pose a significant risk to businesses, threatening security, trust and reputations. Despite research on deepfakes, studies on the risk for companies remain limited. Executives and employees may be unaware of the dangers that deepfake fraud, misinformation and social engineering present. This research seeks to bridge that gap by exploring four key areas: (1) How aware are employees and executives of deepfake threats? (2) How effectively can employees distinguish between real and deepfake images, and what factors influence accuracy? (3) What are the primary risks deepfakes pose to companies? (4) How prepared are companies to detect and respond to deepfake-related threats, and what strategies are most effective?

Using a mixed-method approach combining surveys with employees, and interviews with cybersecurity professionals, this study seeks to enhance corporate awareness of deepfake risks and inform strategies for mitigation. Chapter One introduces the research problem, objectives and methodology. Chapter Two details the methodological framework used, including data collection and analysis techniques. Chapter Three reviews existing literature on deepfake technology and its implications. Chapter Four presents key findings from

primary research, followed by chapter Five which discusses the findings. Finally, chapter Six concludes with a D.E.T.E.C.T Framework recommendation.

2. Methodology

This chapter outlines the methodology used to assess the risk of deepfakes in corporate environments. This dissertation aims to address a gap in existing research concerning the scale, scope, and future implications of this threat. A mixed-methods approach was chosen to enhance the analytical depth of the findings and generate actionable insights for improving mitigation strategies and preparedness (Bachman, et al., 2021; Denscombe, 2021). This chapter explores the research design, data collection, ethics, and the application of analysis.

Research Questions

This study aimed to develop a theoretical framework in response to four research questions (Table 2.1).

Research Questions
RQ1: How aware are employees and executives of deepfake threats?
RQ2: How effectively can individuals distinguish between real and deepfake images, and what factors influence detection accuracy?
RQ3: What are the primary risks deepfakes pose to corporations?
RQ4: How prepared are companies to detect and respond to deepfake-related threats, and what strategies are most effective?

Table 2.1: *Research Questions*

This research employs a cross-sectional mixed methods approach, integrating qualitative and quantitative data to provide both depth and breadth of analysis. Combining numeric data with personal perspectives enhances result clarification, enriches interpretation and offers a nuanced understanding of deepfake risks (Cresswell, 2017; Creswell &

Tashakkori, 2007; Waters, 2019). Data was collected through nine interviews with cybersecurity professionals and a survey of 229 employees. However, mixed methods present challenges when integrating qualitative and quantitative elements with concerns over methodological robustness (Doyle, et al., 2009; Halcomb, 2018; Taherdoost, 2022). However, this approach was deemed most suitable for addressing critical gaps in deepfake research, offering a more comprehensive understanding of the landscape.

Although research is often categorised as either qualitative or quantitative, the distinction between the two is not always rigid, as many studies incorporate both (Hanson, 2008; Maruna, 2010; Newman & Benz, 1998; Ragin, 1994). In this dissertation, criminological and cybersecurity research inherently benefits from a mixed-methods, as both numerical data and qualitative insights contribute to a more holistic understanding of deepfake-related challenges.

Convergent Parallel Design

Triangulation was established in the research community by Webb et al. (1966), arguing that multiple methods established validity and confidence in findings. According to researchers (Denzin, 1978; Patton, 1999), there are four types of mixed methods research designs (Table 2.2).

Research Designs
Triangulation
Embedded,
Explanatory Model
Exploratory

Table 2.2*: Research Designs*

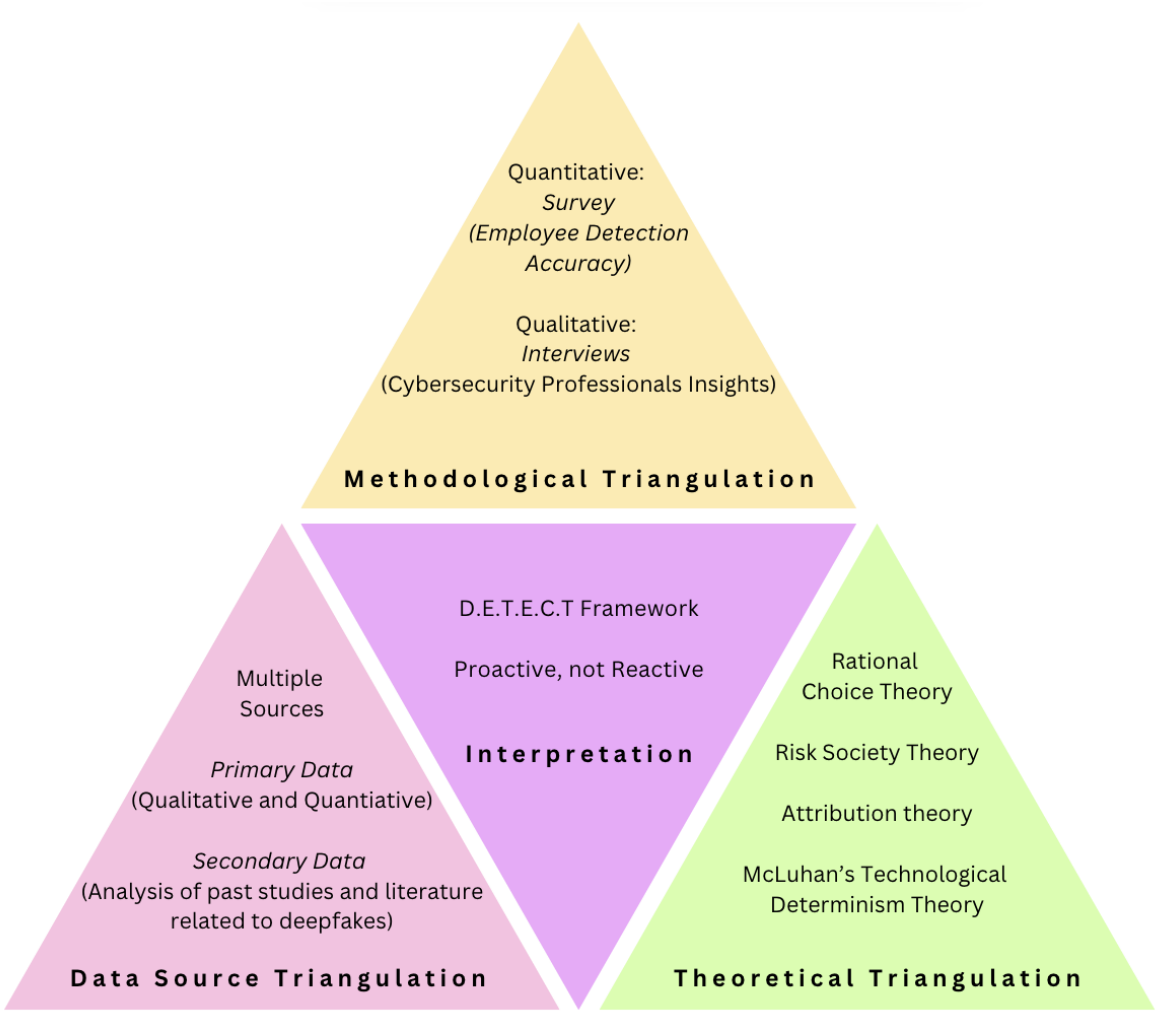


Figure 2.1: *Researchers' Interpretation of Triangulation, adapted from Denzin (1978).*

This study utilises triangulation to corroborate existing literature and gain a comprehensive understanding of the issue (Almalki, 2016; Doyle, et al., 2009; Driscoll, et al., 2007), whilst compensating for the limitations inherent in sole methods, providing a more robust framework for analysis (Figure 2.1). This method was obtained to reduce bias and increase validity whilst obtaining a wider range of perspectives (ibid.). The mixed-methods approach ensures a complete analysis of deepfake risks in corporate environments (Almalki, 2016; Doyle, et al., 2009; Driscoll, et al., 2007). Additionally, secondary research supports primary findings, reinforcing the literature review with established theories and empirical studies (Randolph, 2019; Glass, 1976).

Research Philosophy

This study adopts an inductive approach, exploring the risks of deepfakes in corporate environments through insights from cybersecurity professionals and employee perceptions, rather than testing a pre-existing theory (Bucher, 2021; Thomas, 2006; Thomas, 2003). It is grounded in a constructivist ontology, recognising that perceptions of deepfake risks are shaped by human experiences and evolving technological threats (Cupchik, 2001; Lee, 2012; Packer & Goicoechea, 2000). Epistemologically, the research aligns with interpretivism, seeking to understand subjective experiences through qualitative interviews, while also incorporating positivism elements through a structured survey assessing employees' ability to detect deepfakes (Alharahsheh, 2020; Campbell, 1988; Junjie & Yingxin, 2022; Ryan, 2018).

Secondary research involves collecting and analysing existing data, relying on previously gathered information (Bryman, 2016; Heap & Waters, 2019; Johnston, 2014). It is valuable for comparing data, uncovering trends, confirming theories and providing a broader understanding of the phenomena (Bookstaver, 2021; Johnston, 2014; Sherif, 2018). Additionally, it supports triangulation by complementing primary research and identifying research gaps (Bryman, 2016; Carter, 2014; Hussein, 2009; Oslen, 2004; Walliman, 2021). However, secondary research has limitations, as the data was collected for different purposes, potentially reducing its validity due to outdated or insufficient quality. However, in this study, secondary data was deemed essential when appropriately integrated, enhancing relevance and deepening the understanding of deepfake-related issues (Hox & Boeije, 2005; Johnston, 2014; Young, 2022).

Secondary research revealed a clear gap in the literature. While existing studies examine deepfake advancements, threats, and human ability to detect deepfakes, none explore employees' ability to detect deepfake images in corporate settings. Therefore, this study conducts primary research, analysing employees' ability to detect deepfake content and

how cybersecurity professionals prepare for deepfake threats. Primary research involves gathering new data directly from sources rather than relying on existing studies (Bryman, 2016). This approach provides nuanced insights that can inform companies of the threat, reinforcing the study's inductive and interpretive framework.

However, primary research has limitations. Survey participants may lack deepfake awareness, leading to random guesses. The sample size may limit generalisability, and the rapid evolution of deepfake technology could outdate findings. Additionally, cybersecurity professionals may withhold sensitive information due to confidentiality concerns, limiting the depths of responses. However, primary research remains essential in addressing this research gap.

Data Collection Methods

A pilot study was conducted on the qualitative and quantitative process. For the survey, a small sample (n = 10) of employees tested question clarity and usability, leading to minor revisions and the addition of demographics for deeper analysis. For the interviews, a trial session with one cybersecurity professional helped adjust question phrasing. The pilot also identified technical issues in transcript collection. This process ensured the findings were clear and reliable (Van Teijingen & Hundley, 2001).

Survey

The survey was designed using Microsoft Forms™ and included a mix of images: 18 real images, gathered from online free image websites, and 18 deepfake images, taken from a website that stores GAN Style2 Images (Figure 2.2) (Karras, et al., 2020; Karras, et al., 2024).

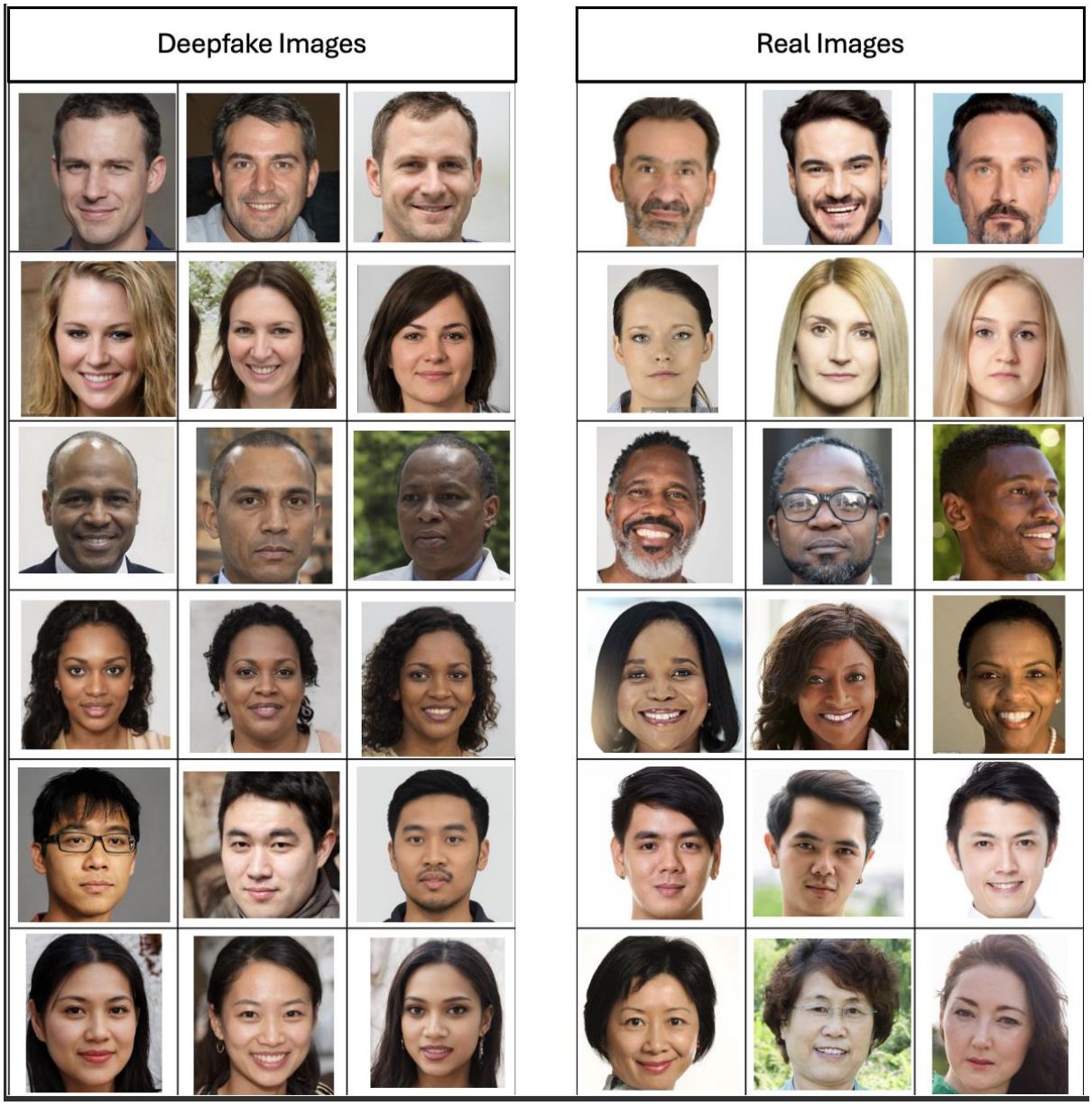


***Figure 2.2**:Deepfake Images and Real Images Used in the Survey*

Participants were shown one image at a time. See Table 2.3 for the required responses. Participants were provided with clear instructions and gave informed consent before proceeding (Appendix A,B).

Information Collected from the Survey	
Image Classification	- Real - Deepfake - Not Sure - Do not want to answer
Confidence Rating	0-10 Scale - (0 being Not at all Confident, and 10 being Very Confident)
Demographics	Age, Gender, Job Hierarchy Level, Industry, Ethnicity, Education Level

***Table 2.3**: Data Collected*

Interviews

Nine semi-structured interviews were conducted with cybersecurity professionals to gain expert insight into deepfake risks in corporate environments. Interviews were recorded and transcribed via Microsoft Teams™, with written participant consent. Each interview lasted 30-60 minutes, during which 22 open-ended questions were asked (Appendix C). Data collection spanned five weeks. Ethical approval was obtained from the university's ethics committee, and all data was anonymised to ensure confidentiality. Semi-structured interviews allowed for in-depth exploration of social phenomena, aligning with the approach of understanding "the social world from the participant's perspectives" (Bryman, 1984, p. 77). Open-ended questions focused on perceptions, experiences, and risks related to deepfakes in corporate settings.

Sampling

Survey

The survey employed a non-probability purposive sampling method to ensure participation from corporate employees, enhancing the relevance of responses (Bachman, et al., 2021). This direct alignment between participants and research objectives strengthened validity (Etikan, et al., 2016; Stratton, 2024), while maintain empathic neutrality (Given, 2008). Initially distributed via LinkedIn™, the survey benefitted from a snowballing sampling effect (Bryman, 2016; Goodman, 1961; Johnson, 2014). However, this method posed limitations. Snowball sampling may have led to homogeneity in perspectives, reducing representativeness (ibid.). Additionally, reliance on LinkedIn excluded employees inactive on professional networking sites, impacting generalisability. Additionally, with the questionnaire being online, this prevented verification of respondent identities, introducing potential sampling bias (Andrade, 2020; Menon & Muraleedharan, 2020; Wright, 2005). However, this strategy was the most effective for reaching professionals with corporate experience. The target sample size was 50 respondents, but 229 responses were collected, significantly improving dataset diversity, robustness and analytical depth, leading to stronger conclusions (Bryman, 2016; Lakens, 2022).

Interviews

The interviewees were selected through a non-probability purposive sampling technique of expert sampling whereby experience and level in the cybersecurity sector were sufficient (Etikan & Bala, 2017; Thapa & Rai, 2015). All interviewees are cyber experts and were recruited through LinkedIn or researcher's connections (Table 2.4). Expert Sampling allowed for high-quality, in-depth participants, which strengthened the findings. Although there is the risk of potential bias with experts having subjective and conflicting opinions, expert sampling was deemed suitable for this research to get a deeper understanding of the deepfake landscape (ibid.). The initial target number of interviewees was three, but nine interviewees were obtained from across the world, generalising validity (Bryman, 2016; Lakens, 2022).

Pseudonymisation will be used to protect participants' identities in this study. Personally identifiable information will be replaced with unique names, ensuring data confidentiality while allowing analysis (Bryman, 2016).

Interviewee	Expertise Area	Location
Interviewee 1, *Irené*	Pre-Sales Lead in Security, bridging technical expertise with business needs.	United Kingdom (London)
Interviewee 2, *Vince*	Cyber Threat Hunting with expertise in Digital Forensics and Malware Analysis.	United Kingdom (London)
Interviewee 3, *Amelia*	Chief Technologist with expertise in networking and security.	United Kingdom (London)
Interviewee 4, *Clara*	Chief Information Security Officer with expertise in cybersecurity.	United Kingdom (London)
Interviewee 5, *Nico*	CEO at a Deepfake Fraud Prevention Firm and Public Speaker.	United Kingdom (London)
Interviewee 6, *Rubén*	Co-Founder at a Deepfake Detection Company.	Italy (Trento)
Interviewee 7, *India*	Co-Founder at a Deepfake Detection Company.	United States of America (New York City) / Israel (Tel Aviv-Yafo)
Interviewee 8, *Amber*	Founder at a Deepfake Detection and Awareness Company.	United States of America (Florida)
Interviewee 9, *Olivia*	3D Artist specialising in State-of-the-Art Deepfake Techniques	Israel (Tel Aviv District)

Table 2.4: *Interviewees*

Analytical Methods

Thematic Analysis

This study employed Braun and Clarke's (2006) six-phase model of thematic analysis, allowing for a rigorous and systematic examination of qualitative data by identifying and analysing themes within the dataset. Data analysis being solely inductive was impossible but recognised as the researcher brings their own pre-conceptual knowledge to the data when analysing. The flexibility of thematic analysis allows exploration of the risk of deepfakes in corporate environments, as it enables the extraction of key concerns, emerging threats, and corporate responses from the participants' perspectives (Figure 2.3) (Braun & Clarke, 2013). To enhance credibility, the accuracy of the findings were strengthened through participant validation, where key insights were discussed with interviewees to ensure their perspectives were accurately represented (Guba & Lincoln, 1994). Additionally, triangulation was used by comparing findings with secondary data sources, such as cybersecurity reports and existing literature on deepfake threats.

***Figure 2.3**: Researchers Interpretation, Braun and Clarke's Six-Phase Model*

Statistical Analysis

Statistical Analysis was used to collect data from the Survey. A combination of descriptive and inferential statistical techniques were employed to examine results and compare them to existing literature. Descriptive Statistics were used to summarise classification accuracy and confidence levels. The mean and standard deviation scores were calculated for correct and incorrect classifications. Visual representations including bar charts, scatter plots, and

heatmaps were used to identify trends and relationships within the dataset. Inferential Statistics were used to determine statistical significance and draw conclusions about the employees (Table 2.6).

Inferential Test	Reasoning
One-Sample T-Test	To compare the mean deepfake detection accuracy to a known population mean (53.16%)
Paired-Samples t-Test	To compare participants' ability to identify real images versus deepfake images
Pearson's Correlation Analysis	To examine the relationship between self-reported confidence and detection accuracy
Factorial ANOVA	To analyse whether demographic factors influenced deepfake detection accuracy
Levene's Test for Homogeneity of Variance	Conducted before ANOVA to confirm variances across different demographic groups were approximately equal

Table 2.5: *Inferential Tests Used*

To ensure rigour, a diverse range of scholarly sources was consulted via Nottingham Trent University's online library and Google Scholar, ensuring credibility (See Table 2.6). Due to fake news, multiple news sources were critically evaluated for accuracy. While literature exists on deepfake technology, research on its direct impact on companies remains scarce, underscoring the need for this study's primary research.

Frequently Searched Words	
"Origins and Development of Deepfakes"	"Generative Adversarial Networks"
"Risk Assessment Frameworks" "Deepfakes"	"Deep Learning"
"Catching cybercriminals"	"Machine Learning"
"Challenges of Catching Cybercriminals"	"Deepfake Creation"
"Deepfakes" "Interviews"	"Deepfake Detection"
"Human Ability" "Deepfakes"	"Financial Fraud" "Deepfakes"
"Deepfake Detection" "Employees"	"Identity Theft" "Deepfakes"
"Corporate Responsibility" "Deepfakes"	"Reputation" "Deepfakes"
"Deepfakes" "Training"	"Opinions of Deepfakes"
"Deepfakes" "Artificial Intelligence"	"UK Policies" "Deepfakes"
"Employee Awareness" "Deepfakes	"Policies" "Deepfakes"
"Trust" "Deepfakes"	"Strategies" "Deepfakes"

Table 2.6: *Frequently Searched Words*

Ethics

Ethics Approval

A professional association's code of ethical practice from The British Society of Criminology was needed for the researcher to conduct the primary research. The researcher was required to complete an ethics form that covered all details about the primary research and prepare key documents (Table 2.7 and Appendix B, C, D, E, F). These materials were submitted to the Ethics Committee for review. The review process took six weeks (Appendix G).

Key Documents for the Ethics Form
Consent Form (Appendix B)
Copy of Interview Schedule Questions (Appendix C)
Participant Information Sheet (Appendix D)
Debrief Form (Appendix E)
Copy of the MS Forms Survey Images (Appendix F)

Table 2.7: *Ethics for Documents*

Informed Consent

Informed Consent was needed from all participants before their involvement in the study. For the survey, the first section required the participants to provide a unique code identifier and read a Participant Information Sheet and confirm their consent before proceeding. For Interview participants, consent forms were distributed in advance, requiring participants to agree to take part. At the beginning of the interview, the researcher read out the study's purposes, participants' rights and the research objectives (Miller, et al., 2012). Participants were informed that they had the right to skip any questions that they were uncomfortable with answering and could withdraw from the study at any time before submission.

Anonymity and Confidentiality

Ensuring anonymity and confidentiality was a priority in this research. For survey participants, anonymity was maintained by requiring them to use a unique identifier code. The researcher's email was provided at the beginning and end of the survey, allowing participants to withdraw their responses if desired. However, it was acknowledged that complete anonymity could not be fully guaranteed if participants chose to withdraw after submission (Bryman, 2016; Clark, et al., 2021). For Interview participants, the consent

form also required a unique identifier code, ensuring that no personally identifiable information was linked to their responses. Many interviewees were willing to use their name and company, although anonymity was decided on. While the researcher and participants understood that absolute anonymity could not be guaranteed, confidentiality was maintained by storing all data on a password-protected laptop accessible only to the researcher.

3. Literature Review

Deepfake technology is advancing at an unprecedented speed, leaving companies ill-equipped and vulnerable to attacks. This literature review examines existing research on the risks associated with deepfakes and their implications for companies. A literature review is essential for synthesising prior research, identifying trends, and highlighting knowledge gaps (Denney & Tewksbury, 2013; Machi & McEvoy, 2022).

Definitions and Evolution of Deepfakes

Deepfakes are AI-generated synthetic media designed to mimic real individuals in videos, images or audio recordings. Deepfakes emerged in 2017 on Reddit™ and the term was coined from "Deep Learning" and "Fakes" (Burkell & Gosse, 2019; Kietzmann, et al., 2020; van der Nagel, 2020; Westerlund, 2019), emphasising technology's ability to create hyper-realistic, yet entirely fabricated content. Unlike editing techniques, including Photoshop, deepfakes leverage artificial neural networks to produce highly detailed, adaptive, and realistic outputs (Goodfellow, et al., 2014; Goodfellow, et al., 2020; Karras, et al., 2020; Karras, et al., 2024). Deepfake technology can be beneficial in multiple cases (Table 3.1) (Chesney & Citron, 2019; Farid, 2022; Westerlund, 2019; Usukhbayar & Homer, 2020).

Category	Benefits
Film and TV (Kietzmann, et al., 2020)	- Fix misspoken lines via voice dubbing - Modify Scripts - Multilingual adaptation
Gaming (Headshot, 2025)	- Create 3D characters
Actors and Stunt Doubles (European Parliament, 2021)	- Use digital actors for new footage - Enhance realism - Alter actor ages
Historical (Kerner & Risse, 2020)	- Recreate deceased individuals - Re-enact historical events
Video Conferencing (European Parliament, 2021)	- Life-life avatars - Improve video quality - Reduce bandwidth with facial animation - Direct eye contact
Entertainment via Social Media (AI Video, 2025; Apex Heroes, 2025; Battleitout, 2025; Beyond The Screen, 2025; CEOshrek, 2025; Daddy Shark, 2025; Deranged AI, 2025; DreamWeaver_AI, 2025; Hatim's Shorts, 2025; MEGA, 2025)	- Fake videos of celebrities, that users find entertaining (See Figure 3.1 and Appendix H.)

Table 3.1: *Benefits of Deepfakes*

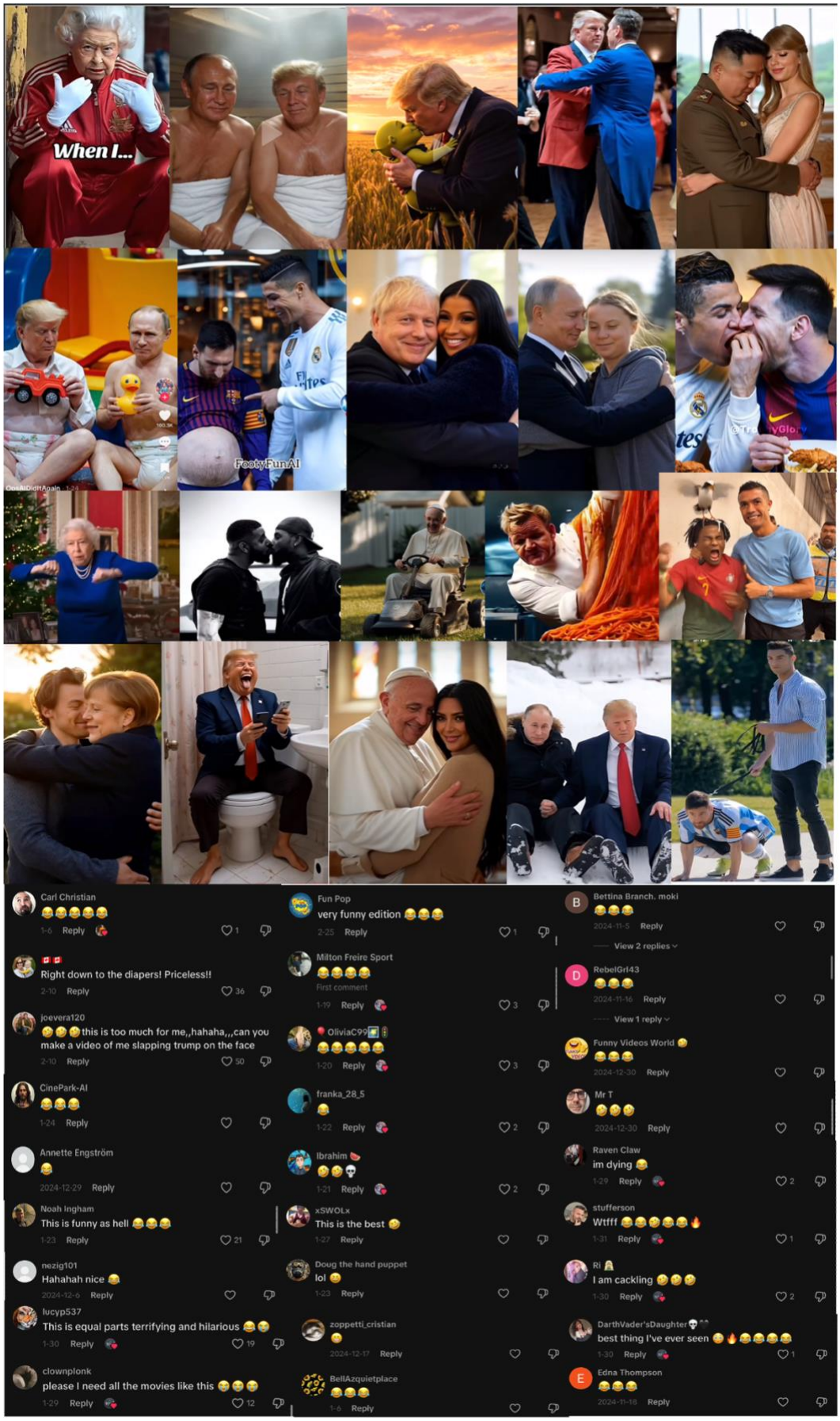


Figure 3.1: *Entertainment via Social Media (Also Appendix H)*

However, the proliferation of deepfake technology has given rise to significant security and ethical concerns. Entrust (2024) and SumSub (2024) found that 2023 was the first year that deepfakes became a widespread attack vector (Figure 3.2), attacks occur every 5 minutes and deepfakes overall have increased by 4x (Figure 3.3). Furthermore, Schreiber and Schreiber (2024), suggest significant knowledge gaps regarding AI risks among employees, with a majority unaware of deepfake threats.

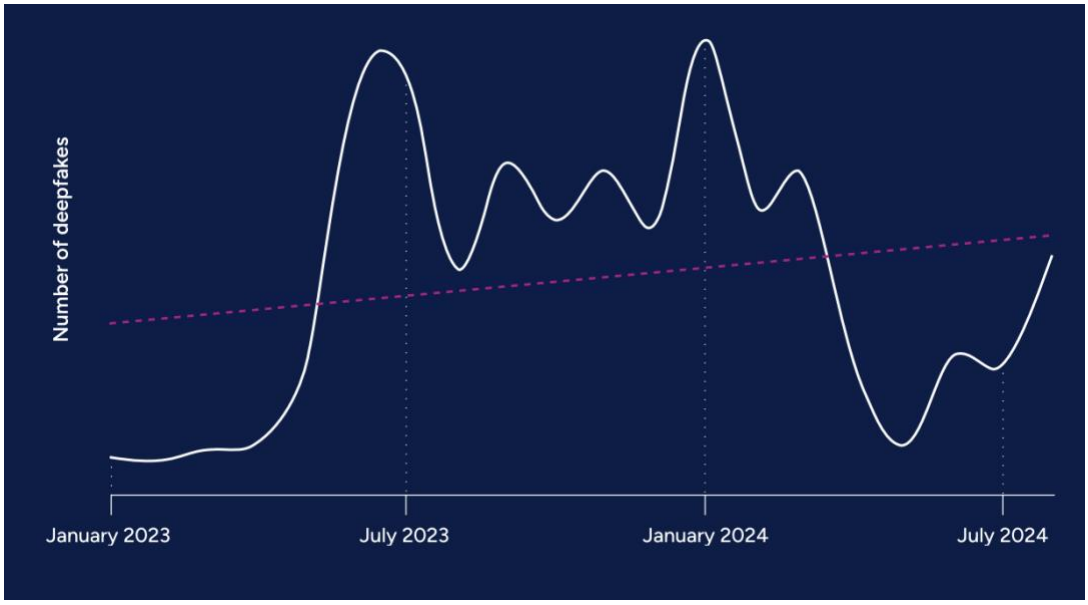


Figure 3.2: *Number of Deepfakes between January 2023 - July 2024*

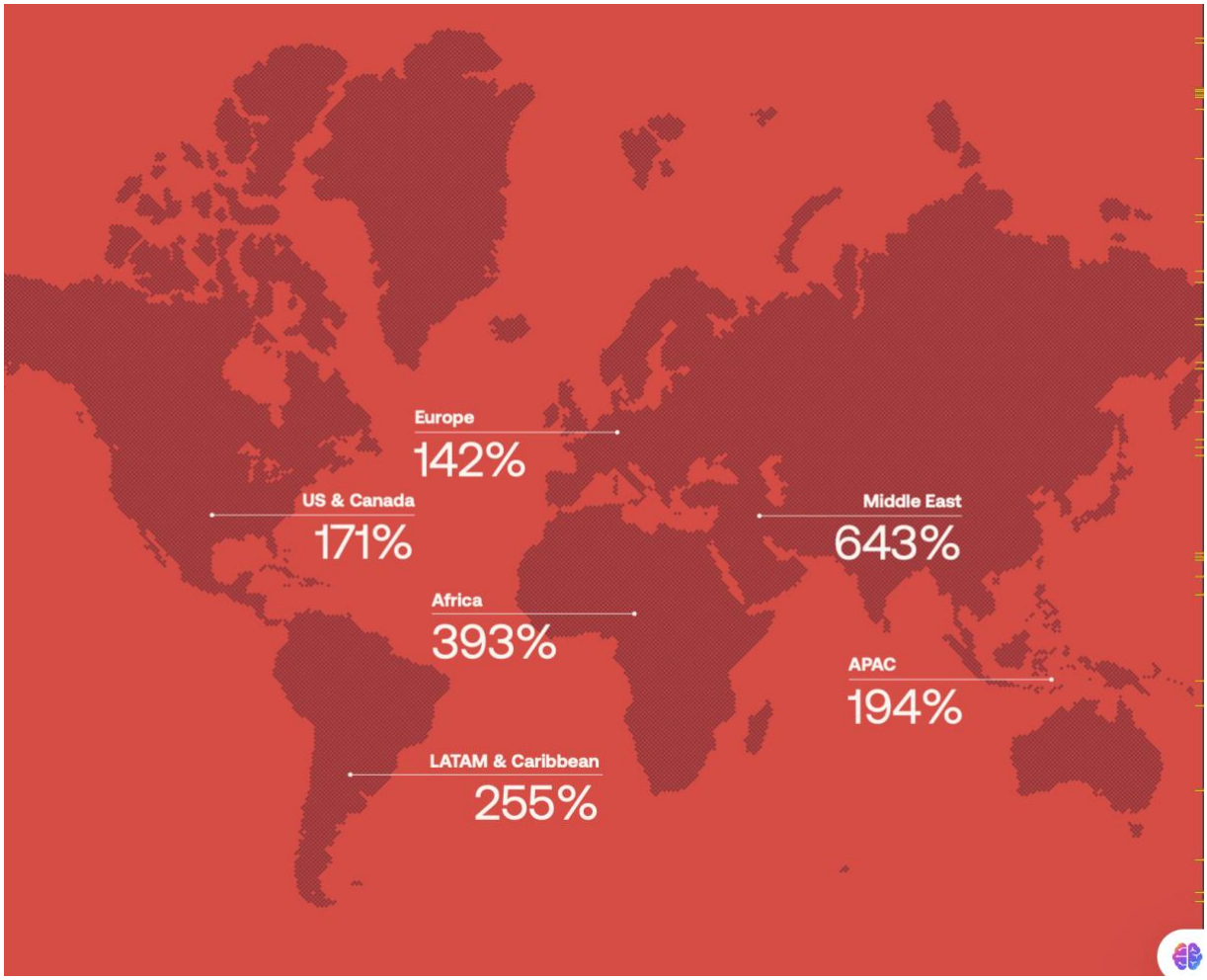


Figure 3.3: *Average deepfake fraud growth by region, 2024*

Deepfake technology has been marked by significant milestones in AI development. Deep faking efforts started in the 1990s, where AI-generated media relied on basic vision models, which lacked realism and required manual input (Bregler, et al., 2023; Masood, et al., 2023; Tolosana, et al., 2020). Early facial animation and voice synthesis tools struggled with the "uncanny valley" effect, making faces easily detectable (Tinwell, 2014). A groundbreaking discovery was made in 2014 by Ian Goodfellow, unveiling Deep Learning and GANs (Goodfellow, et al., 2014; Goodfellow, et al., 2016; Goodfellow, et al., 2020).

The threat of Deepfake technology to companies comes with deep learning technology, which simulates a human brain by taking large data sets and learning from them (Holdsworth & Scapicchio, 2024; Sze, et al., 2017). Brain-inspired computing, or neuromorphic computing, supports this dynamic by creating systems that mimic the

brain's neural architecture (Mehonic, et al., 2020; Mehonic & Kenyon, 2022). Open-source AI tools such as 'StyleGan2' and 'DeepFaceLab' have lowered the barrier for non-experts to create realistic deepfake content, with over 95% of deepfake videos created on DeepFaceLab (Github, 2024), intensifying concerns over corporate security (Karras, et al., 2020; Perov, et al., 2020; Widder, et al., 2022). These incidents underscore the criminological significance of deepfakes, which enable forms of impersonation, fraud, and sabotage that challenge traditional corporate security measures. Analysing deepfake-related risks from a criminological perspective is crucial in explaining how and why offenders exploit this technology and necessary preventative measures.

Overhyped or a Legitimate Threat?

While some scholars argue that deepfake risks are overstated (Acerbi, et al., 2022; Altay, et al., 2023; Ecker, et al., 2022; Mercier, 2020; Simon, et al., 2023), rapid advancements in AI-driven forgery technology necessitate urgent corporate action. Critics suggest that the high costs of generating ultra-realistic deepfakes make them an unlikely tool for mass corporate fraud, with some deepfake creators charging £20,000 per minute of synthetic video (Figure 3.4) (Kaspersky Daily, 2023; Kaspersky Threat Intelligence Portal, 2024).

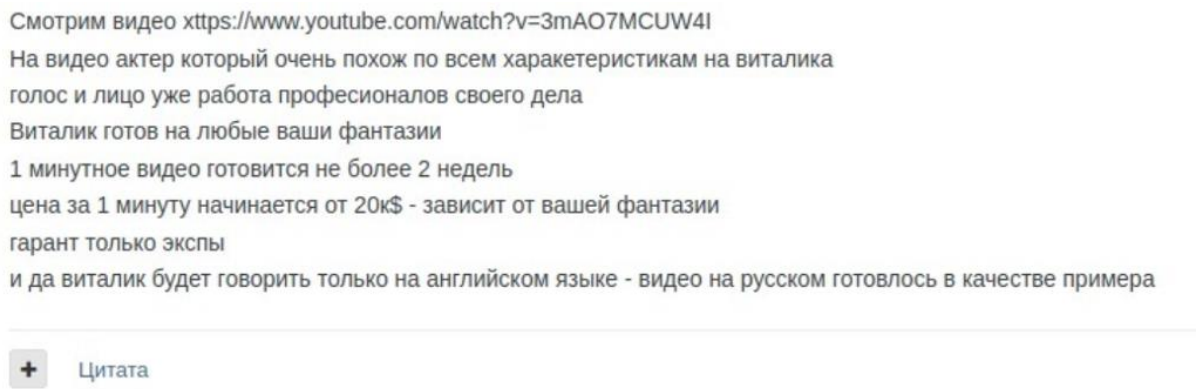


Figure 3.4: *Dark Web Offer*

Counterarguments highlight that deepfake technology is becoming more accessible and difficult to detect, increasing corporate risk (Karras, et al., 2020; Perov, et al., 2020; Widder, et al., 2022). A key challenge of deepfake risks lies in corporate complacency. Studies show that only 29% of companies have taken action against deepfakes, despite evidence that the fraud landscape shows deepfakes having a strong frequency, volume and high threat level, and account for 40% of biometric fraud (Entrust, 2024). Furthermore, 17% of employees take an action that leads to significant financial loss (Breacher.ai, 2025). This discrepancy raises concerns about whether corporations are underestimating the technological shift that deepfakes represent.

Deloitte (2024) found 25.9% of executives say their companies have experienced one or more deepfake incidents. Deepfake attacks in companies focus on identity fraud, impersonating individuals to gain access to data or money (Entrust, 2024; Gordon & Ma, 2003). Between 2022 and 2024, deepfake-related cybercrime surged by 3000%, correlating with increased accessibility of open-source deepfake software and AI-driven identity forgery services (Entrust, 2024; Karras, et al., 2020; Perov, et al., 2020; Widder, et al., 2022).

Rapidly Evolving Landscape

McLuhan's (1975) Technological Determinism Theory argues that technology itself shapes societal behaviours and institutions. Under deepfake technology, one could argue that the rapid development of AI-driven media forces corporations to react rather than proactively shape security strategies. With the speed of advancements, corporations may not have time to put strategies in place before becoming a target (Chesney & Citron, 2019; Ikenga & Nwador, 2024). If companies fail to adapt to AI threats, they are at risk of inevitable vulnerabilities as technology drives change regardless of human intervention (Brynjolfsson & McAfee, 2014; O'Neil, 2016). This aligns with neuromorphic computing, where AI

systems mimic human cognition, enabling autonomous adaptation and reducing the extent of human control over digital threats (Mehonic, et al., 2020; Mehonic & Kenyon, 2022).

This perspective assumes that corporate executives lack preparation in shaping security policies, which may not be the case (Borrett, et al., 2014; Wilshusen & Powner, 2009). Companies may prioritise other cyber risks over deepfake threats, dedicating fewer resources to deepfake-specific policies while focusing on higher-risk areas (Eling, et al., 2021; Entrust, 2024). This aligns with Ulrich Beck's (1992) Risk Society Theory in suggesting that modern societies are increasingly governed by risk management strategies. In this context, corporations must integrate AI-driven threat mitigation into their security policies, ensuring cyber-risk prevention becomes a core business function rather than merely an IT concern (Bojanc & Jerman-Blažič, 2008).

Alternatively, corporations are not merely passive responders to technological change but can actively shape the trajectory of deepfake risks (Benaroch, 2018; Sun, et al., 2023; Wang, et al., 2015; Spears & Barki, 2010). By investing in cybersecurity and advocacy for AI regulation, companies can influence how deepfake threats evolve, demonstrating that while technology drives change, human intervention remains a critical factor in shaping its impact (ibid.).

Deepfake Generation

So how do Deepfakes work? Deepfakes operate by leveraging sophisticated deep-learning models to manipulate and synthesise hyper-realistic content. These systems rely on neural networks trained on extensive datasets of a target individual's facial expressions, voice patterns, and mannerisms (Güera & Delp, 2018; Rana, et al., 2022; Seow, et al., 2022; Zhang, 2022). By analysing and mapping intricate facial movements, deepfake algorithms can seamlessly superimpose a subject's likeness onto another person's face, creating convincing yet synthetic representations.

Advanced facial landmark tracking and motion synthesis techniques ensure deepfake content remains realistic, aligning with head movements, lighting, and micro-expressions, making detection difficult (Karras, et al., 2020; Kietzmann, et al., 2020; Whittaker, et al., 2023). While these advancements drive innovation, they pose serious risks, particularly in misinformation, identity fraud, and manipulation (Korarkar & Sakarkar, 2023). As detection struggles to keep pace, researchers are developing forensic tools to analyse inconsistencies in pixel distribution, temporal coherence and biological signals distinguishing real footage from synthetic fabrications (Demir & Ciftci, 2021; Güera & Delp, 2018; Rombach, et al., 2022; Seow, et al., 2022; Van Den Oord, et al., 2016; Zheng, et al., 2021). There are three general deepfake creation models (Table 3.2).

Main types of Deepfake Creation Models	
Autoregressive model	Predicts the next value in a sequence using previous values (Seow, et al., 2022; Van Den Oord, et al., 2016; Rombach, et al., 2022)
Autoencoder	Compresses and Reconstructs data via an encoder-decoder structure (Güera & Delp, 2018; Seow, et al., 2022).
Generative Adversarial Networks	Two neural networks compete to create realistic synthetic data (Creswell, et al., 2018; Goodfellow, et al., 2014; Goodfellow, et al., 2016; Goodfellow, et al., 2020; Gui, et al., 2021; Hermosilla, et al., 2021; Karras, et al., 2019; Karras, et al., 2020).

Table 3.2: *Main Deepfake Models*

Generative Adversarial Networks (GANs)

While easily accessible deepfake generators may not always achieve the highest level of realism, their increasing availability still raises significant concerns. At the core of deepfake generators lies GANs, a powerful machine learning framework that has significantly advanced the realism and sophistication of synthetic media. This dissertation assumes GANs as the primary enabler of deepfake technology, supported by primary research involving the use of deepfake images using GAN-based software. GANs operate through a dual-network system and once trained, GANs can produce convincing synthetic images from randomly sampled latent spaces (Table 3.3 and Figure 3.5) (Creswell, et al., 2018; Goodfellow, et al., 2014; Goodfellow, et al., 2016; Goodfellow, et al., 2020; Gui, et al., 2021; Hermosilla, et al., 2021; Karras, et al., 2019; Karras, et al., 2020).

How Generative Adversarial Networks Operate	
The Generator	Creates synthetic data that mimics real data. It takes random input vectors and creates data samples to fool the discriminator into believing the generated data is real.
The Discriminator	Tries to distinguish between real and fake data by taking the data as input and classifying them as real or fake. It then tries to improve its accuracy over time.

Table 3.3: *The generator and the discriminator*

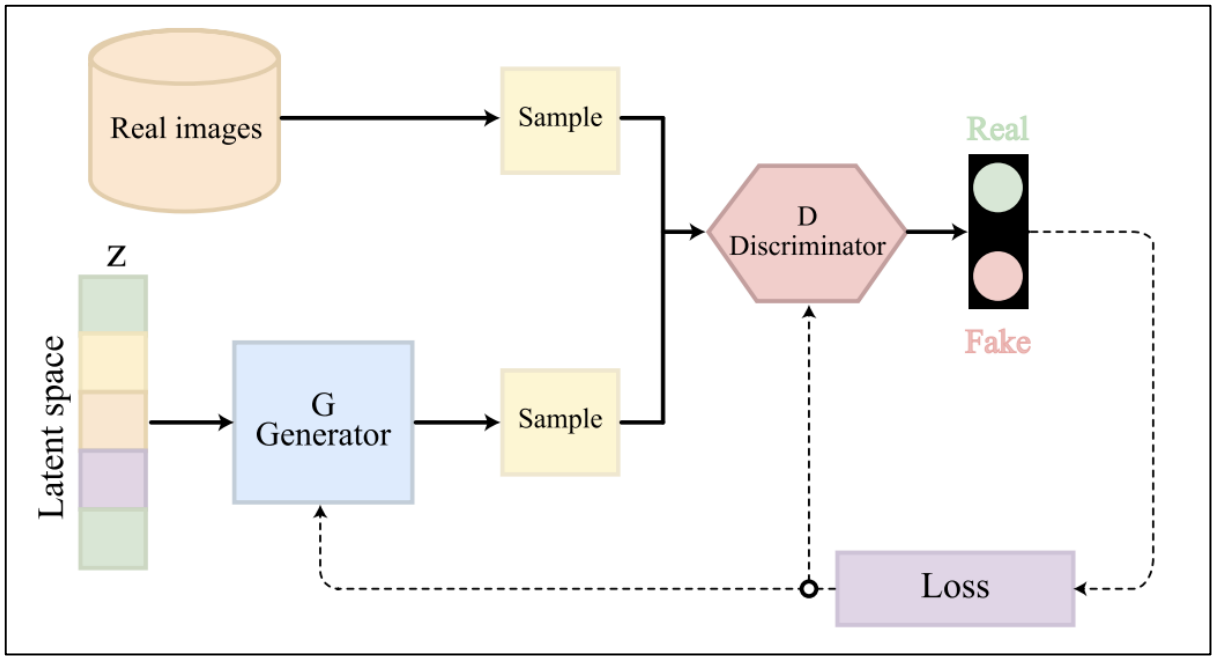


Figure 3.5: *GAN Diagram*

(Karras, et al., 2024; Tolosana, et al., 2020), developed by Nvidia researchers, represents the most efficient synthetic images (Figure 3.6 and Table 3.4). StyleGAN generates high-resolution synthetic images with remarkable photorealism (Karras, et al., 2020; Karras, et al., 2024; Karras, et al., 2020).

Types of Facial Manipulation
Entire Face Synthesis
Attribute Manipulation
Identity Swap
Expression Swap

Table 3.4: *How GANs work*

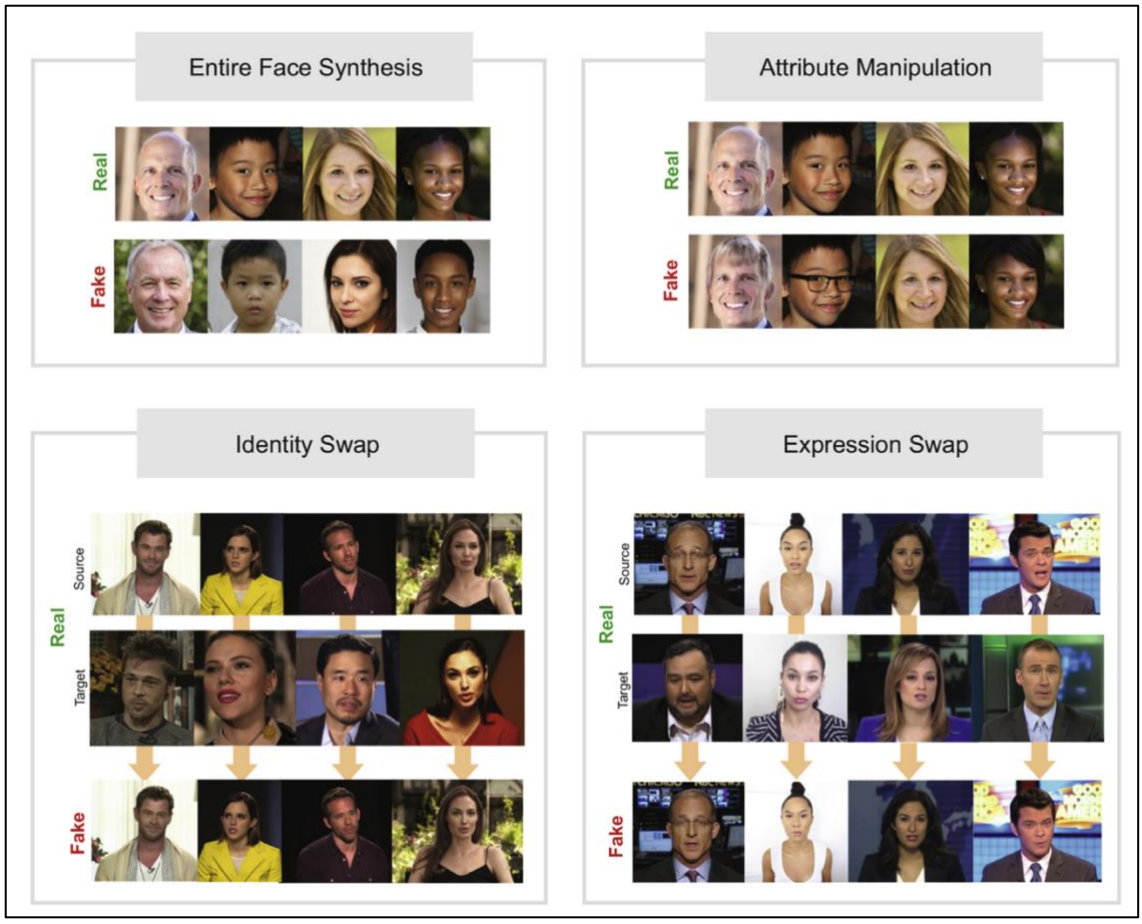


Figure 3.6: *Types of Facial Manipulation*

Past Studies

Studies on human deepfake detection abilities (images) are scarce (Lyu, 2020). Over 56 studies have been conducted to understand the human ability to detect deepfakes, including image, video and audio experiments (Diel, et al., 2024a). To the researcher's knowledge, there are no studies that assess how well employees in companies can detect deepfakes, but there are 12 studies on the human ability to detect deepfake images (Table 3.5), which is the closest correlation to this study's research. These studies report accuracy rates between 44% and 75%, suggesting that humans struggle to identify deepfakes. While some findings indicate that training and AI assistance can improve accuracy (Boyd, et al., 2023; Cartella, et al., 2024; Diel, et al., 2024b; Hulzebosch, et al., 2020; Kramer & Cartledge, 2024; Robertson, et al., 2018), others highlight persistent overconfidence

(Bray, et al., 2023; Nichols, et al., 2022; Tucciarelli, et al., 2022). The limited reliability of human detection underscores the potential risks for businesses. Looking at all studies on deepfake detection (image/ video/ audio), Ahmed and Chua (2023) found that ethnicity did not affect detection ability, but accuracy improves when participant demographics align with those of the deepfake images (Khan, et al., 2023). El Mokamem (2023) and Tahir et al. (2021) found that training assisted in detection accuracy.

Author	Year	Sample Size	Age	Participant Type	Type of Deepfake	Accuracy Rate	Incentives	Key Findings
Boyd, et al.	2023	N = 1,560		Random	Image	65.70%	Money	Humans correctly identified real faces 2/3 of the time but struggled with synthetic ones. AI assistance improves accuracy
Bray, et al.	2023	N = 280	M=26	Random	Image	62%	Money	Participants deteced deepfake better than chance but remained inaccurate, overly confident and unaffected by simple interventions.
Cartella, et al.	2024	N = 20			Image			Sematic-aware edits were the hardest to detect, whilst instruction-guided edits were the easiest to identify as fake
Diel, et al.	2024	N = 96	M = 33	Random	Image	44.22%		Training impreoved accuracy by 20% (44.22% to 64.74%), but increased distress and AI anxiety. Awareness of detection inability reduced self-efficacy.
Hulzebosch et al.	2020	N = 496		Random	Image	65.30%		Humans correctly identified CNN generated images 70.1% of the time, influenced by factors like AI experience, resolutioin and feedback
Kramer & Cartledge	2024	N = 260	M = 28.6	Convenience	Image	57.90%		Using crowd based methods significantly improves human detection of deepfakes, with the wisdom of outer crowds yielding better accuracy compared to individual performance
Liu, et al.	2020	N = 20			Image	72.72%		Deepfake faces exhibit distinct texture differences from real faces and Gram-Net architecture significantly improves the robustness and generalisation ability to detect deepfakes
Nichols, et al.	2022	N = 227			Image	75.43%, 62.92%		Professional experience had no clear impact, and some experts performed worse. Error rates were high but some excelled (96.30% accuracy)
Nightingale & Farid	2022	1. (N = 315), 2. (N = 201)		Convenience	Image	48.2%, 59%	Money	AI-Synthesises faces are now indistinguishable from real ones and deemed more trustowrthy, surpassing previous limitations in photorealism.
Robertson, et al.	2018	N = 80	M = 30		Image	48%		Training significantly improved morph detection rates from 48% to 79%. Detection accuracy correlates with face matching skills. Training benefits low performers more than high performers
Shen, et al.	2021	1. (N = 176), 2. (N = 174), 3. (N = 172)	19-69 years	Convenience	Image	49.1%, 49.7%	Money	Human participants were generally unable to distingush between synthetic and real faces. Different backgrounds did not really make an impact.
Tucciarelli, et al.	2022	N = 107	28.18	Random	Image			Participants judged deepfkae faces as more "real" than actual real faces leading to increased confidence in their judgements

Table 3.5: *Past studies on human deepfake detection abilities (Images)*

Theoretical Framework

Rational Choice Theory (RCT) provides a useful lens to examine the decision-making processes of both deepfake perpetrators and corporate entities responding to these

threats. As originally formulated by Becker (1968), RCT posits that individuals use a cost-benefit analysis when making decisions, weighing potential rewards against associated risks. Recent studies have extended this framework to cybercrime, illustrating that cybercriminals assess the likelihood of detection, legal repercussions, and financial gain before executing the attack (Holt & Bossler, 2015). Applying this to deepfakes, Mandelcorn et al. (2013) and Brewer et al. (2019) suggest that malicious actors exploit technological advancements and regulatory gaps to optimise their strategies, thereby maximising profit while minimising exposure (Figure 3.7). Computer crime can be seen as profitable and low risk. Motivations can be determined by financial gains, including fund theft, selling stolen data or leveraging trade secrets, or non-financial gains, including revenge, hate crime, challenge and blackmail (ibid.).

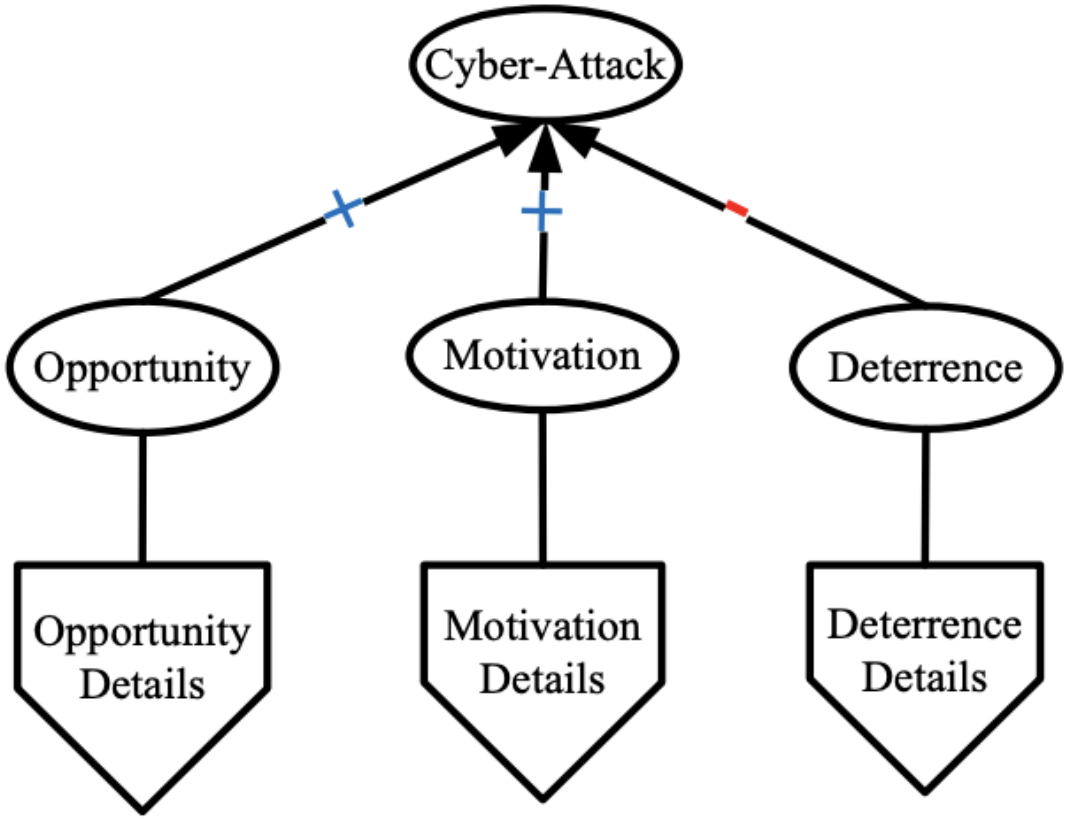


Figure 3.7: *Motivations for Cyberattacks*

Opportunity suggests cyberattacks can be planned or random (Figure 3.8). In corporate settings however, the application of RCT suggests that organisations may fail to prioritise deepfake mitigation due to a perceived low probability of risks, including immediate financial loss or reputational damage (Alahmari & Duncan, 2020; Eling, et al., 2021; Entrust, 2024).

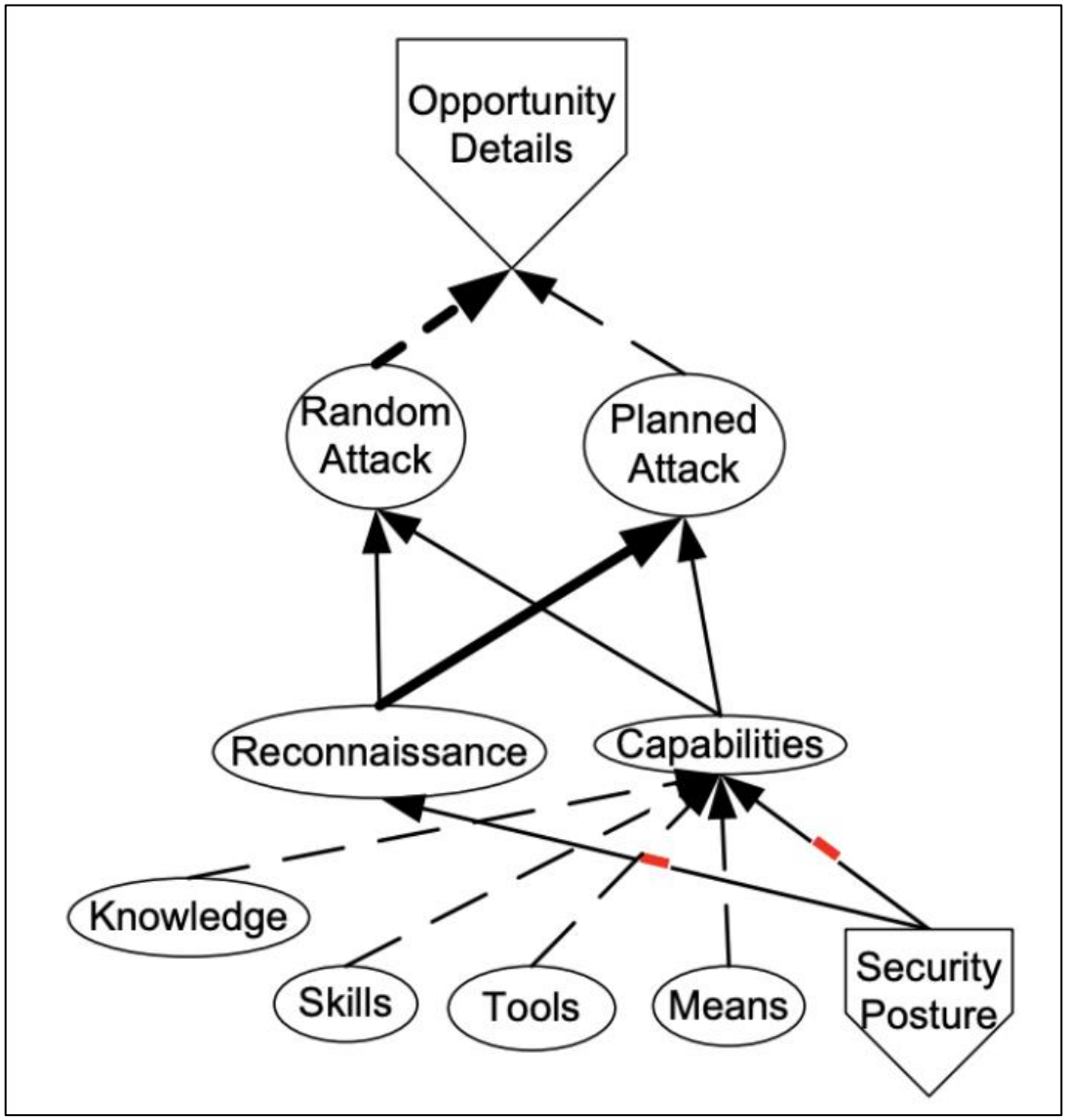


Figure 3.8: *Opportunity for cybercriminals to attack*

Risks and Threats of Deepfakes to Corporations

Deepfake technology poses a threat to corporate environments through deception and manipulation (Table 3.6) (ADL, 2023; Agarwal, et al., 2019; Bateman, 2022; Flick & Morehouse, 2011; Miller, 2021; Munk, 2024; Salahdine & Kaabouch, 2019; SumSub, 2024). Fraudulent AI-generated videos and audio have been used to deceive executives and employees into transferring millions through video meetings. Fraudsters used deepfakes to impersonate Arup's Chief Financial Officer and other employees during a video call. As a result, an unsuspecting staff member executed 15 financial transactions, transferring nearly $26 million to fraudulent bank accounts in Hong Kong (Leng & Ho-him, 2024; Magramo, 2024; Milmo, 2024; Noto, 2024; Smith, 2024; World Economic Forum, 2024). Between March 2024 and 2025, 7 deepfake attacks have been reported by companies (Appendix I), and 198 total deepfake reports in total across the world (Resemble.AI, 2025) (Appendix J). With the rise of remote working, cybercriminals are using deepfake technology to bypass online interviews (Figure 3.9). This incident highlights the urgent need for robust deepfake detection mechanisms and heightened cybersecurity measures.

Psychological Harm	Financial Harm	Societal Harm
- Defamation - Intimidation - Bullying - Undermining Trust	- Extortion - Identity Theft - Fraud - Stock-price manipulation - Brand damage - Reputational damage	- News media manipulation - Damage to economic stability - Erosion of trust

Table 3.6: *Researchers' adaptation the European Parliaments (2021) outline of deepfake risks to companies*

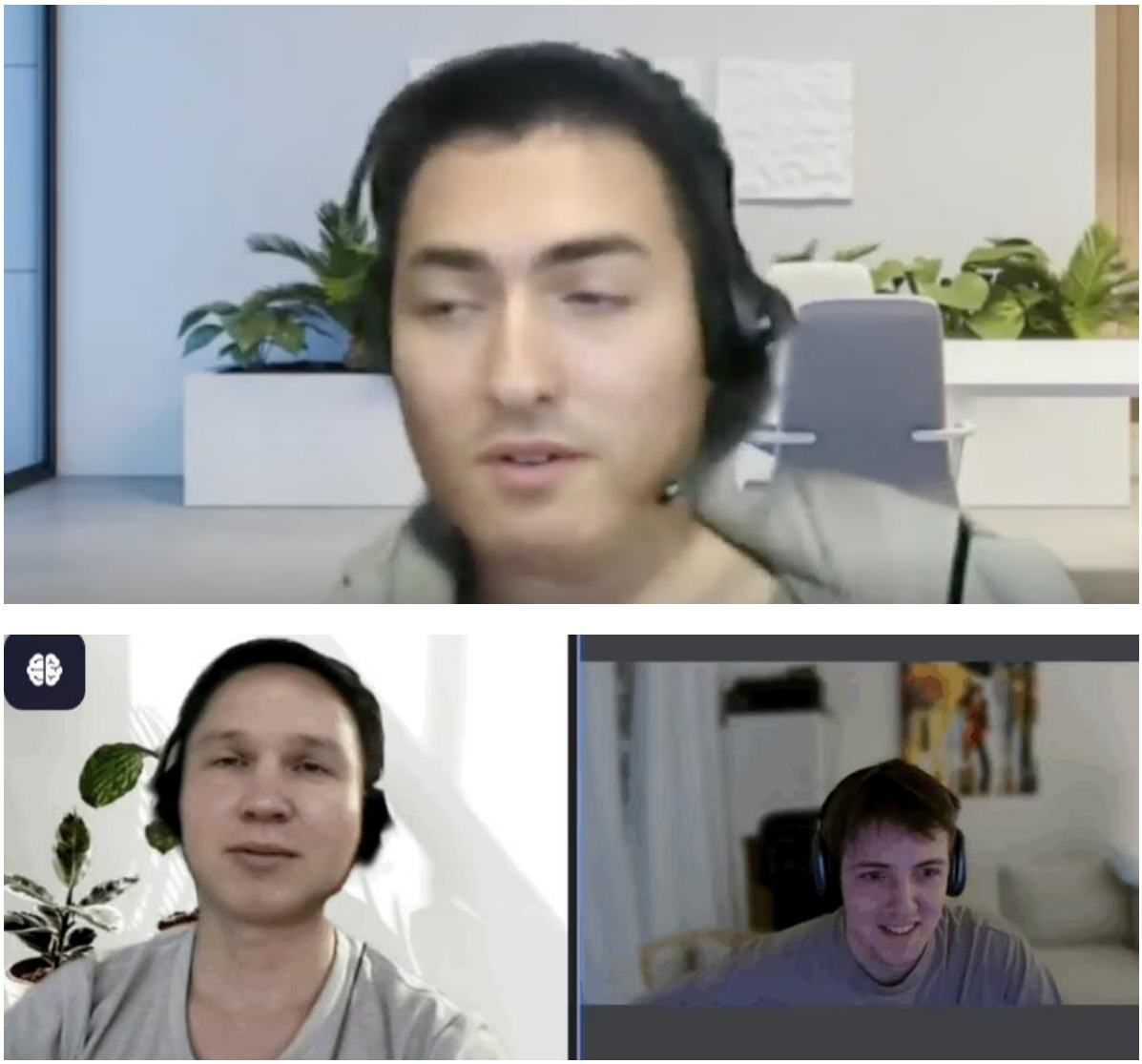


Figure 3.9: *Real Interviews with individuals using deepfake technology (Liporazzi, 2025; Moczadlo, 2025).*

This rapid expansion has transformed deepfakes from a niche innovation into a global security and misinformation crisis. Deepfakes are now widely exploited for fraud, disinformation, and cybercrime, posing significant geopolitical, economic, and social threats (Entrust, 2024; Miller, 2021; Sharma & Kaur, 2022; Westerlund, 2019). In China, a cybercriminal employed real-time deepfake video technology to impersonate a victim's acquaintance during a live video call. As a result, the victim was deceived into transferring over $600,000, demonstrating the capabilities of deepfake technology (Global Times, 2023; Samson, 2023; Saxena, 2023; Simonchik, 2025; Zekun, 2023).

Cybercrime-as-a-service has evolved into a complex, highly organised service, where sophisticated tools and malicious services are readily available to a broad range of users through online marketplaces (Manky, 2013). Deepfakes have emerged as critical cybersecurity, enabling fraud, social engineering, identity theft, and reputational manipulation. As AI advances, deepfakes-as-a-service are becoming more advanced, posing significant risks to individuals, corporations and national security (Europol, 2022; Gamage, et al., 2022). Criminals sell tools, technology and knowledge to facilitate cybercrimes, outpacing law enforcement in executing and adapting new tools (Europol, 2022).

Identity Theft Fraud

The increasing sophistication of deepfake technology has introduced unprecedented security vulnerabilities, particularly in biometric authentication and multi-factor identification verification. Traditionally, biometric security has been considered a highly secure safeguard against unauthorised access (Ghilom & Latifi, 2024; Jain & Kumar, 2012; Millett & Pato, 2010). However, synthetic media advancements have significantly undermined these security measures, allowing cybercriminals to impersonate individuals with alarming precision, leading to unauthorised transactions, corporate espionage, and data breaches (Flick & Morehouse, 2011). With just one photo and three seconds of audio, a deepfake can be generated (Bontcheva, et al., 2024; Kietzmann, et al., 2020; Microsoft, 2025a; Wang, et al., 2023). Images can be made easily on applications (Castillo Camancho & Wang , 2021; Tolosana, et al., 2020), such as the AI App 'Zao' (Antoniou, 2019; Coleman, 2019; De Seta, 2021; Doffman, 2019; France-Presse, 2019). Furthermore, 'Krea.ai' allows users to generate images depicting someone's identity by taking a selfie alongside a synthetic driver's license (Figure 3.10), further escalating biometric risk. As multi-factor authentication (MFA) becomes standard, incorporating passwords, PINs, secondary devices, facial and fingerprint recognition (Bhargav-Spantzel, et al., 2006; Kim

& Hong, 2011; Microsoft, 2025b; Ometov, et al., 2018), companies must critically reassess the conventional identity verification systems to mitigate deepfake-driven fraud.

Figure 3.10: *'Krea.ai' allows users to create synthetic images with ID photos – Prompt: "Create an image of a human holding an ID card" (Krea.AI, 2025).*

A study by Entrust (2024) highlighted that deepfakes are the new face of Video Biometric Fraud, accounting for 40.8% of attacking mechanisms (Figure 3.11). Furthermore, deepfake videos have also been used to subvert 'Know Your Customer (KYC)' identification verification procedures, which are widely employed in financial sectors and cryptocurrency exchanges (ibid.)

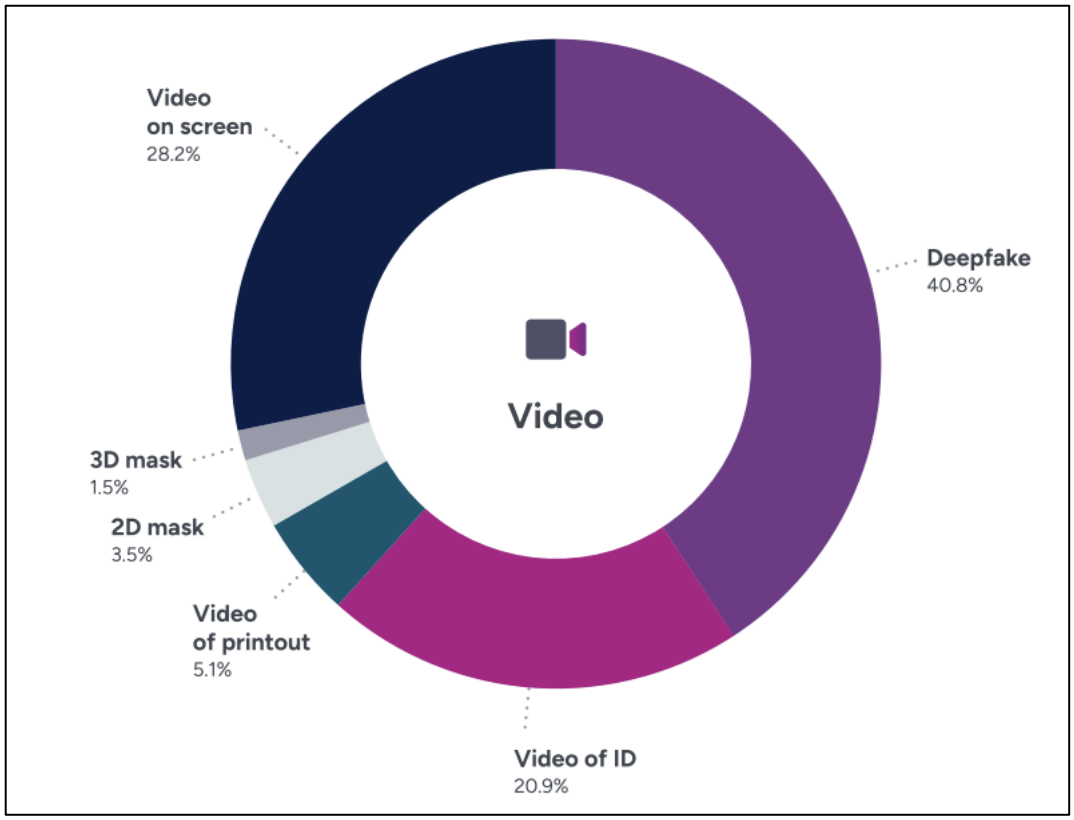


Figure 3.11: *Deepfakes account for 40.7% of Biometric Fraud*

These instances highlight the weaknesses of biometric security against deepfake-driven deception. While MFA was developed to enhance security (Bhargav-Spantzel, et al., 2006; Kim & Hong, 2011; Microsoft, 2025b; Ometov, et al., 2018), deepfake technology enables cybercriminals to circumvent advanced verification protocols. Vulnerabilities lies in the reliance on static authentication. Standard facial recognition systems, lacking liveness, can be fooled by deepfake images and videos (Agarwal & Farid, 2021; Entrust, 2024). Similarly, voice authentication systems fail to distinguish between real and synthetic speech patterns, rendering them susceptible to deepfake-generated manipulations (Almutairi & Elgibreen, 2022; Lim, et al., 2022; Wang, et al., 2020). AI-powered deepfake detection tools' effectiveness remains limited as adversaries continue to refine their techniques to evade detection (Yu, et al., 2021; Wang, et al., 2024, p. 25). Agarwal et al. (2020) advocates for a more robust approach to address vulnerabilities and deter cybercriminals. However, Bhargav-Spantzel et al. (2006) argue that MFA remains an

adequate safeguard when biometric security is supplemented with additional verification layers.

While some companies are actively investing in AI-driven deepfake detection and security frameworks, the pace at which deepfake technology is evolving often exceeds the ability of companies to implement countermeasures effectively (Entrust, 2024). This raises concerns about the long-term sustainability of biometric authentication and a viable security solution (Karras, et al., 2020; Perov, et al., 2020; Widder, et al., 2022).

Social Engineering

Social engineering Attacks (SEAs) constitute a significant cybersecurity risk, as they exploit human psychology to manipulate individuals into divulging sensitive information or granting unauthorised access (Dsouza, et al., 2024; Peltier, 2006, p13; Syaritri, et al., 2022). Despite advancements in cybersecurity, human vulnerability remains a fundamental weakness as humans find it difficult distinguishing between true and false information (Marsh, et al., 2016; Ncubukezi, 2022; O'Connor & Weatherall, 2019; Salahdine & Kaabouch, 2019; Sanders, et al., 2019). As Schultz (2005) posits, security threats are inherent to human-managed systems, as they are susceptible to errors and manipulation. The introduction of deepfake technology has exacerbated these risks, facilitating highly realistic impersonation techniques that increase the effectiveness of SEAs.

SEAs typically follow a four-phrase process: research, trust-building, exploitation and exit (Dsouza, et al., 2024; Krombholz, et al., 2015). Traditionally attackers relied on deceptive emails, phone calls or physical impersonation. However, deepfakes have revolutionised social engineering by enabling cybercriminals to fabricate highly convincing video and audio content. Flick & Morehouse (2011) outline methods in which deepfakes enhance

SEAs (Table 3.7). These strategies capitalise on trust and urgency, both of which are psychological levers frequently exploited in SEAs.

Social Engineering Attack Methods
- Impersonating senior executives to authorise fraudulent financial transactions
- Mimicking IT personnel to obtain login credentials
- Generating deceptive video calls to manipulate employees into sharing confidential data

Table 3.7: *SEA methods*

The effectiveness of deepfake-enhanced SEAs lies in their ability to subvert traditional verification methods. As Geddes (2020) and Tahir et al. (2021) observe, the phrase "seeing is believing" is no longer a reliable standard for authentication, necessitating a paradigm shift in cybersecurity strategies, despite mundane visual communication (Costa, et al., 2020). Furthermore, the debate persists as to whether deepfake risks stem primarily from technological advancements or human susceptibility to manipulation (Ncubukezi, 2022). A particularly salient real-world example is the case of KnowBe4 (2024), where a North Korean spy used AI-enhanced deepfake technology to fabricate an identity and successfully pass the hiring process at a cybersecurity firm. This incident underscores the emergence of deepfake-driven insider threats, demonstrating that even companies specialising in cybersecurity are at risk.

Reputational Damage

Unlike conventional reputational crises, deepfake-related attacks blur the boundary between fact and fiction, making it difficult for companies to control their public image (Geddes, 2020; Tahir, et al., 2021). Given the rapid dissemination of misinformation via social media and news platforms, deepfake-induced crises can destabilise investor

confidence, influence stock prices, and erode brand credibility by fabricating damaging content (Al-Khazraji, et al., 2023; Bateman, 2022; de Rancourt-Raymond & Smaili, 2023; Kalaiarasu, et al., 2024; Petratos, 2021; Singh & Dhiman, 2023; Vosoughi, et al., 2018).

Attribution Theory suggests individuals assign responsibility for negative events, shaping their perception of corporate credibility (Heider, 1958; Jones, et al., 1972; Weiner, 1974; Weiner, 1986). Deepfake misinformation can have immediate and measurable consequences (Vâlsan, et al., 2022), as seen in 2024, when a deepfake video of Ashishumar Chauhan, CEO of the National Stock Exchange (NSE), recommending stocks went viral (National Stock Exchange of India, 2024). Although NSE debunked the video, the incident exposed investor vulnerability. Gwebu et al. (2018) warn that corporate silence can worsen stock price fluctuations, regulatory scrutiny, and reputational damage. Say and Vasudeva (2020) reported that firms remain vulnerable due to internal systemic issues, while Foerderer and Schuetz (2022) found stock markets relatively unresponsive. Shim and Yang (2016) and Sjovall and Talk (2004) argue that delayed or concealed responses lead to greater reputational consequences. Once consumer trust is compromised, customer loyalty declines (Ahmad, et al., 2023; Ahmed & Chua, 2023; Huang & Maracic, 2024; Whittaker, et al., 2021). Furthermore, public scepticism towards corporate statements is rising (Grier & Forehand, 2003; Walker, 2005). The case of Arup illustrates the effectiveness of transparent crisis management, reinforcing Vecchietti et al.'s (2025) argument that companies must proactively manage deepfake crises whilst ensuring prompt, clear communication (Knight & Nurse, 2020). This proactive crisis management enabled Arup to control the narrative, dispel misinformation, and maintain stakeholder trust (Wahyu, 2023).

Despite examples of effective crisis response, deepfake-related incidents often go unreported (Gregory, 2025). Lydon (2021) and Sikra et al. (2023) suggest corporations may conceal deepfake-driven fraud to protect reputation. While this may offer short-term risk mitigation, unaddressed deepfake threats erode digital trust and expose organisations

to future cyberattacks. Companies are often reactive, rather than proactive. Kwon and Johnson (2014) suggest security investments are made after an attack reinforcing a reactive security reaction, rather than proactive. Say and Vasudeva (2020) argue that signs of attacks may suggest underlying vulnerabilities in the company's security systems. Borrett et al. (2014) and Wilshusen and Powner (2009) support this in stating that organisations are falling behind in the fast-paced threat environment.

Corporate Preparedness and Response Strategies

Regulatory Compliance and Legal Limitations

Currently, there is no comprehensive legislation in the United Kingdom (UK) specifically focused on deepfakes attacking companies. While existing laws such as the Fraud Act (2006), Data Protection Act (2018), Defamation Act (2013), and Intellectual Property Act (2014) could potentially be adapted to address deepfake-related threats, their effectiveness remains inadequate. GDPR offers some protection for victims (Moreno, 2024), but UK laws currently focus primarily on sexually explicit deepfakes under the Sexual Offences Act 2003, following amendments to the Online Safety Act 2023 (Ministry of Justice & Davis-Jones MP, 2025), leaving gaps in protection for corporate entities affected by deepfakes.

A key challenge in relying on legal measures is that identifying and prosecuting deepfake cybercriminals is difficult, (Brenner, 2012; Das & Nayak, 2013; Hartono, et al., 2024). Some corporations have implemented their own proactive measures rather than relying solely on legislative protections. Both public and private companies can employ risk assessment frameworks to evaluate their exposure to deepfake threats. Public companies may be more likely to invest in AI-driven detection technologies due to greater financial resources and regulatory expectations (Table 3.8). Furthermore, deepfake detection

companies have appeared in the last couple of years (Table 3.9). Despite advancements, detection remains a persistent challenge. The continuous development of GANs allows deepfake creators to evade existing detection algorithms (Yu, et al., 2021; Wang, et al., 2024, p. 25). Additionally, dataset manipulation and video compression techniques make it more difficult to identify fraudulent content (Europol, 2022).

Deepfake Detection Technologies
Biological signal analysis (Ciftci, et al., 2020)
Phoneme-viseme mismatch detection (Agarwal, et al., 2020)
Facial movement tracking (Liao, et al., 2023)
Recurrent Convolutional Models (Sabir, et al., 2019)

Table 3.8: *Detection Software*

Company	Main Goal
'Honor' - Chinese smartphone brand	Continuous frame-by-frame monitoring to analyse discrepancies in eye contact, lighting, image clarity, and video playback (Honor, 2024).
TrueBees	Authenticate online images by exploiting the synergy between digital media forensics and blockchain technology (TrueBees, 2025)
Reality Defender	Analyses photos and videos at pixel level to determine authenticity, and real-time voice detection to assess truth (Reality Defender, 2025).
Clarity	Uses real-time AI threat detection to identify digital impersonations (Clarity, 2025).
Sensity	System examines pixels, file structures and voice patterns (Sensity, 2025).
Truepic	Computer vision and blockchain technology to verity the authenticity of photos and videos when they are taken (Truepic, 2025).
Breacher.ai	Offer services including deepfake phishing simulations, custom training modules and advanced detection tools (Breacher.ai, 2025).
Loti	Scans the internet every day for traces of target individual (e.g. Executive), then takes down infringing accounts and content (Loti, 2025).
Shreem Growth Partners	Deepfake fraud prevention consulting firm, equipping businesses with tools, knowledge and strategies (Shreem Growth Partners, 2025).
Trusona	Account Take Over Protection, using multiple authoritative data sources, assesses risk factor (Trusona, 2025).
GetReal Labs	Forensic tools detect voice and video deepfakes in real-time, and authentication technology ensures integrity (GetReal Labs, 2025).
Paravision AI	Facial Recognition, Liveness Detection, Deepfake detection and Age Estimation (Paravision AI, 2025).

Table 3.9: *Upcoming Deepfake Detection Companies*

While deepfake technology poses a growing risk, policymakers have avoided implementing overly restrictive regulations that could withhold AI advancements (GOV.UK, 2023; Kyle, 2025). This approach reflects a broader strategy to position the UK as a global leader in AI development, highlighting a preference for supporting AI-driven solutions to counter deepfakes, rather than reinforcing restrictive measures. Public Companies are more likely to engage in collaborations with government agencies and technology firms to address the deepfake threat. The UK Government Certified Programme (IASME, 2025), and the National Cyber Security Centre's Framework (NCSC, 2025) provide structured approaches for managing security risks, detecting cyber threats and minimising impacts. Furthermore, the AI Safety Institute is allocating a total of £8.5 million in grants to researchers working on AI-based security measures, promoting the development of corporate defence strategies against deepfake threats (Kyle, et al., 2024). However, some countries seem to be more prepared (Table 3.10).

Country	Legislation
India	**Use of existing Laws**
	Information Technology Act (2000)
	- Section 66 (C) for identity theft
	- Section 66 (D) for cheating or fraud
	Indian Penal Code
	- Section 500: Defamation (including deepfake-based defamation)
	- Section 468 – Forgery using deepfakes
	- Section 124 – Sedition or inciting hatred via deepfakes
	- Section 506 – Threats or intimidation using fabricated content
China	**Regulations on the Management of Deep Integration of Internet Information Services (2023)**
	- Article 6: Prohibits deepfake content that violates laws, including fake news
	- Article 7: Requires service providers to implement security and algorithm assessments
	- Article 9: Mandates real identity verification for users

	- Article 10: Obligates providers to review and manage synthesised content - Article 11: Requires mechanisms for debunking false information - Article 14: Enforces strict data and biometric information protection - Article 116: Requires labelling of AI-generated or edited content
United States of America	**Federal Legislation** - Deepfakes Accountability Act (H.R. 5586, 2023): Requires AI-generated content disclosure, penalises fraudulent deepfakes - Protecting Americans from Foreign Adversary Controlled Applications Act (2024): Regulates AI misinformation impacting corporations - NDAA 2021: Directs Homeland security to assess deepfake threats to national security and businesses **State Laws** - California AB 730 (2019): Bans deepfakes in elections - Texas SB 751 (2019): Criminalises deepfake misuse in political and public influence - Mississippi SB 2577 (2024): criminal penalties for wrongful dissemination of deepfakes
European Union	**EU AI Act** - Risk-based Classifications of AI systems - Compliance and Penalties - Expected to become a global AI regulation benchmark - Digital Services Act (DSA)

Table 3.10: *Deepfake Legislation in other countries*

Training and Awareness

Training and awareness programmes provide companies with the necessary expertise and abilities to recognise and mitigate cyber risks (Munk, 2015). Private companies still seek to mitigate risks through education and training. This includes raising awareness among clients and employees to help them recognise credible information sources (Mustak, et al., 2023). Some researchers argue that resistance employee training programmes can

significantly mitigate SEAs, suggesting that the human factor remains the most critical point of intervention (Aldawood & Skinner, 2019, p. 73; Ghafir, et al., 2016; Salahdine & Kaabouch, 2019). Furthermore, studies have suggested training and education methods to reduce likelihood of a deepfake attack (Chi, et al., 2020; Naffi, et al., 2025; Solo, 2025), and although focused predominantly on youths, these strategies can be applied to employees. Bhalli et al (2024) found that training students significantly improved deepfake detection.

4. Findings

Descriptive Statistics - Survey

This section outlines the descriptive statistics of the study. The research sample comprised 229 participants (90 Female, 137 Male, and 2 Prefer not to say) through Microsoft Forms, using a non-purposive sampling strategy. Although LinkedIn users' average age is between 25-34 (Bondar, 2023; Dixon, 2024), this study managed to target a range of ages between 20-70 (M_{age} = 42.79, SD_{age} = 14.34). Furthermore, demographics of Ethnicity, Education Level, Job Hierarchy and Industry were collected to observe whether there was any correlation between demographics and ability to detect deepfakes (Table 4.1 and Figure 4.1).

Gender	Female	Male	Prefer not to say
Total	90	137	2

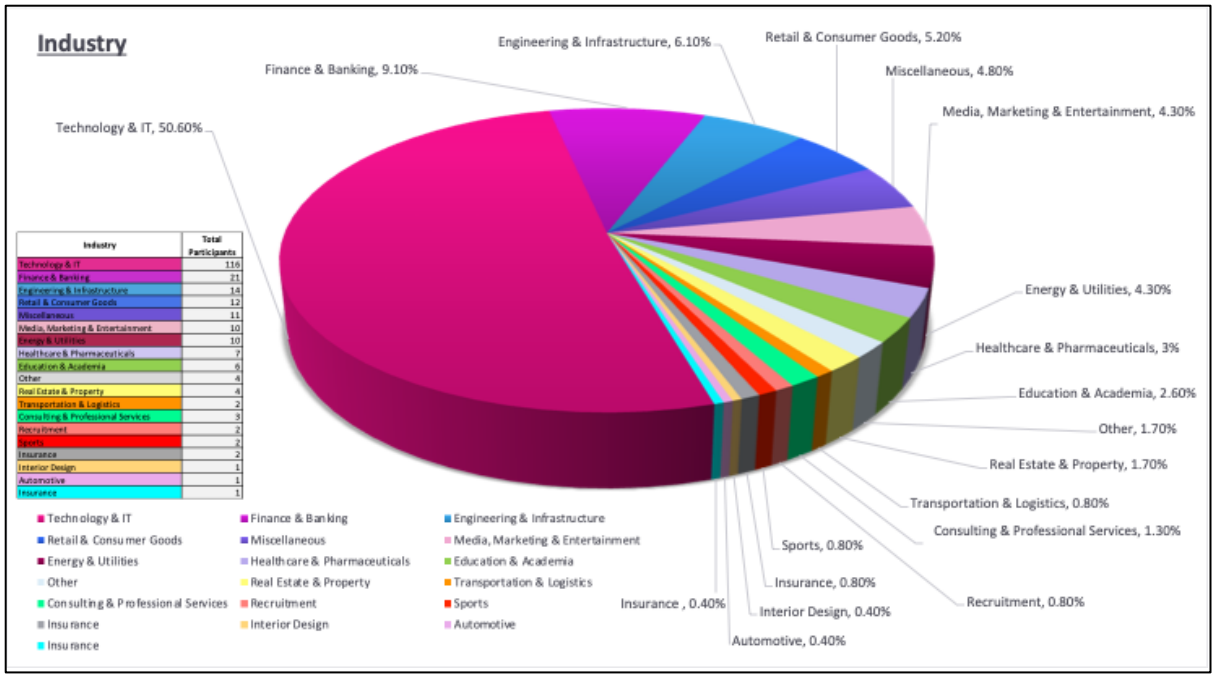
Age	Gen Z	Millennials	Gen X	Baby Boomer	*Mean*	*SD*
Total	61	48	97	23	42.79	14.34

Ethnicity	Black	White	Asian	Mixed	Other
Total	12	190	12	9	6

Education Level	High School	BTEC	A/O Levels	Undergraduate Degree	Postgraduate Degree	Other
Total	12	17	32	114	38	16

Job Level	Board of Directors	Executive	Director	Manager	General Employee	Entry Level	Other
Total	19	19	38	48	70	21	14

Table 4.1: *Statistics by Demographics*



Industry	Total Participants
Technology & IT	116
Finance & Banking	21
Engineering & Infrastructure	14
Retail & Consumer Goods	12
Miscellaneous	11
Media, Marketing & Entertainment	10
Energy & Utilities	10
Healthcare & Pharmaceuticals	7
Education & Academia	6
Other	4
Real Estate & Property	4
Transportation & Logistics	2
Consulting & Professional Services	3
Recruitment	2
Sports	2
Insurance	2
Interior Design	1
Automotive	1
Insurance	1

Figure 4.1: *Industry of Participants*

Statistical Analysis

This section of research sets out the findings of the survey on deepfake detection accuracy among employees and executives in response to Research Question 2 (Table 4.2).

RQ2: *How effectively can individuals distinguish between real and deepfake images, and what factors influence detection accuracy?*

Table 4.2: *Research Question 2*

The data is explored in five key areas: overall accuracy rates, demographic rates, common errors, confidence levels, and comparison to expectations using SPSS software. The survey took on average 10.51 minutes to complete (SD=15.56 minutes).

The participants were first asked to fill in details about their demographics of age, gender, ethnicity, education, job hierarchy and industry. The participants were shown one image

at a time, and were asked whether they believed the image was real, deepfake, not sure, or do not want to answer. After each image, they rated their confidence in their answer on a Likert scale of 0 (not at all confident) to 10 (very confident). This allowed us to measure confidence per image.

A one-sample t-test was conducted to compare employees' deepfake detection accuracy to a hypothesised population mean of 53.16, derived from 56 previous studies (Diel, et al., 2024a). This benchmark was selected as it represents the expected level of performance observed in multiple studies, providing a valid comparison for assessing participants' ability to distinguish between real and deepfake content. The test examined whether participants' accuracy in this study was significantly different from the established standards. Results showed that the mean accuracy in this sample (M=32.36%, SD=11.33, N=229) was significantly lower than the hypothesised benchmark, $t(228) = -27.77$, $p<.001$. The mean difference of -20.80% (95% CI: -22.27, -19.32) indicates a substantial decline in detection performance. The effect size was large (Cohen's D =1.83), suggesting a major difference between corporate employees' detection ability and prior benchmarks, highlighting a critical vulnerability in deepfake recognition within corporate settings (Appendix K).

Analysing the videos individually reveals that, for only 2 out of the 36 images, participants' guesses were more accurate than in previous studies (Figure 4.2).

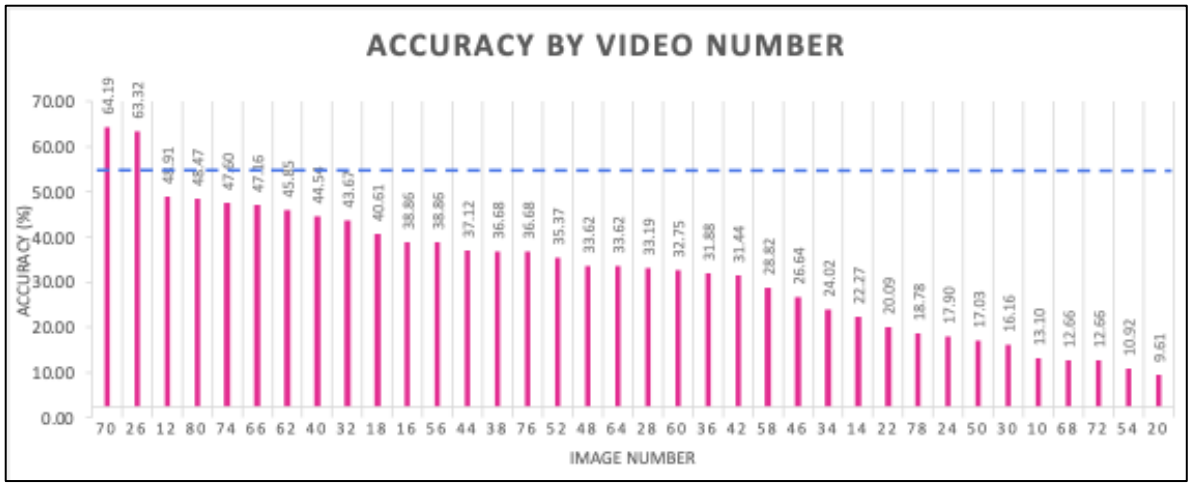


Figure 4.2: *Accuracy by Video Number*

The blue line indicates the average hypothesised mean (Diel, et al., 2024a).

Looking at detection accuracy by fake vs. real provides a deeper insight into how well participants guessed. For real and deepfake images, only one image correctly identified more than the average hypothesised mean (Figure 4.3). Participants' accuracy in correctly identifying deepfake images was 30.16% (SD = 16.95), while their accuracy in correctly identifying real images was 34.57% (SD = 17.32). A paired-samples t-test showed that participants were significantly more accurate in identifying deepfake images (M= 34.57%, SD=17.32%) compared to real images (M=30.16%, SD=16.95%), t(228)=2.60, p=.010. The mean difference was 4.41% (95% CI: 1.07, 7.76), indicating that the observed difference is statistically different (Appendix L). Furthermore, there was a difference of 52% accuracy rates between the highest and lowest correctly identified deepfake images (Figure 4.4).

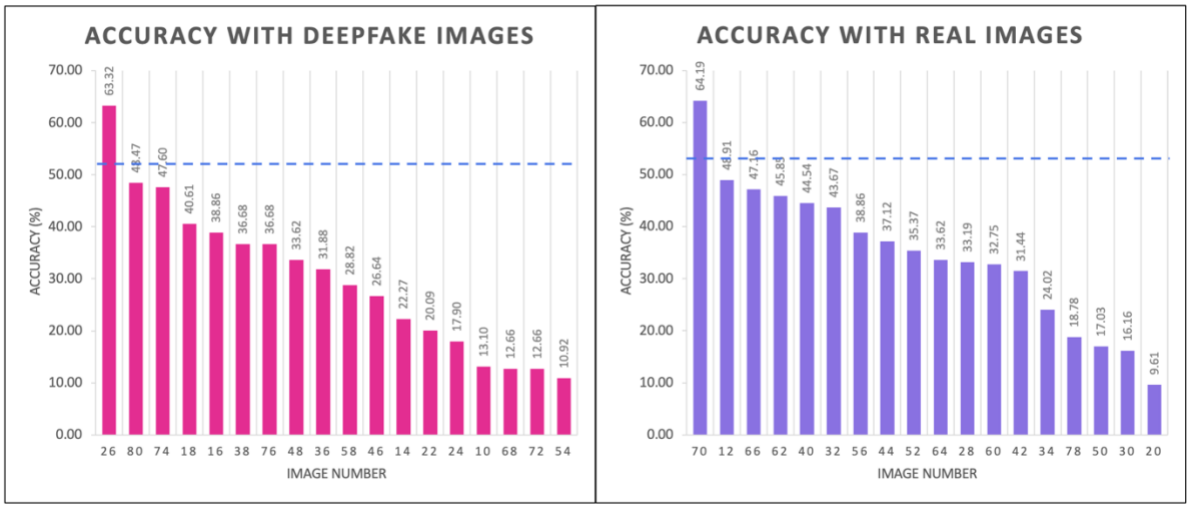


Figure 4.3: *Accuracy scores between real and deepfake images.*

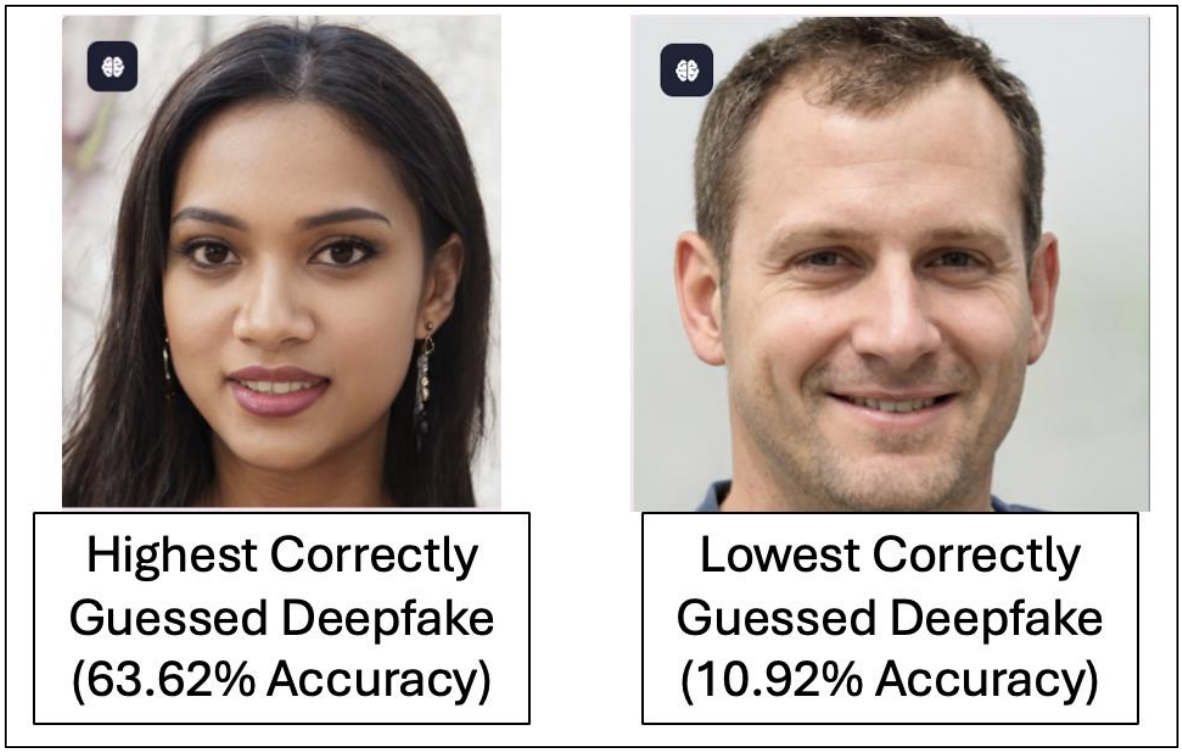


Figure 4.4: *Highest and Lowest Correctly Guessed Deepfake Images.*

Analysis of the self-reported confidence to detect deepfakes revealed an average confidence level of 6.35 (SD = 0.214) on the 10-point scale. A Pearson's correlation analysis was conducted to examine the relationship between confidence levels and deepfake detection accuracy. The results showed a very weak negative correlation, r(229) = -0.101, *p = .129.* This suggests that confidence levels were not significantly associated

with accuracy in deepfake detection. Since the correlation was non-significant, there is no strong evidence that individuals with higher confidence performed better or worse in detecting deepfakes (Figure 4.5 and Table 4.3).

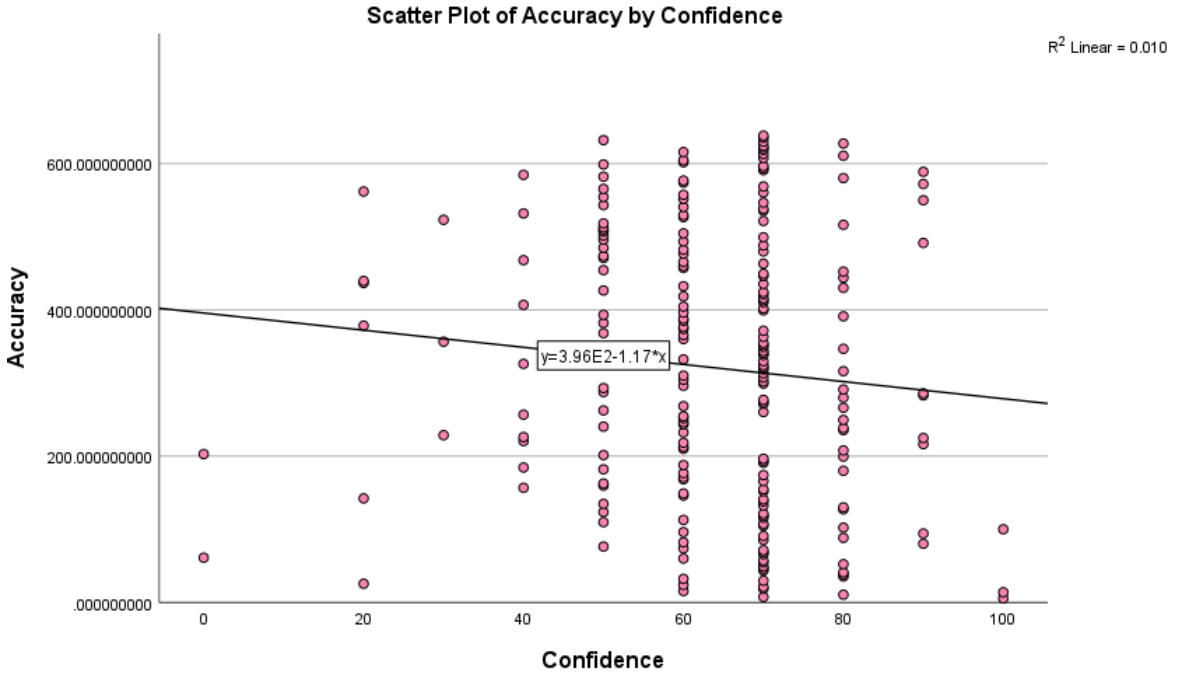


***Figure 4.5**: Pearsons Scatter Plot of Accuracy by Confidence*

		Confidence	Accuracy
Confidence	Pearson Correlation	1	-.101
	Sig. *2-tailed		.129
	N	229	229
Accuracy	Pearson Correlation	-.1-1	1
	Sig. *2-tailed	.129	
	N	229	229

***Table 4.3:** Pearsons Correlation*

A Factorial ANOVA was conducted to examine the effects of age, ethnicity, education, rob level, industry and gender on deepfake detection accuracy. Levene's test for homogeneity of variance was non-significant, F(178,50)= 0.684, p=.962, indicating that the assumption of equal variances was met. This suggests that the differences between group variances are unlikely to affect the interpretation of the ANOVA results. The analysis revealed no significant main effects for any of the independent variables, suggesting employee demographics did not significantly influence deepfake detection ability (Table 4.4).

Demographic	F(df1, df2)	p-value	Partial η² (Effect Size)
Age	F(3, 50) = 0.517	.671	.014
Ethnicity	F(4, 50) = 0.703	.592	.025
Education	F(5, 50) = 1.044	.395	.046
Job Role	F(6, 50) = 1.104	.361	.034
Industry	F(17, 50) = 1.064	.392	.087
Gender	F(2, 50) = 1.015	.364	.011

Table 4.4: *Factorial ANOVA*

A heatmap was generated to visualise the mean deepfake detection accuracy across different demographics, as some participants did better than others (Figure 4.6).

Average Correct		Black (M)	Black (F)	White (M)	White (F)	Asian (M)	Asian (F)	Mixed (M)	Mixed (F)	Average Correct by Employee Level
Employee Level	Entry Level	38.89		35.30	36.67			36.11		36.74
	General Employee	38.89	25	33.32	31.04	25	38.89		61.11	36.18
	Manager	25		34.29	29.92	34.72	55.56	23.61		33.85
	Director			27.93	34.07	27.78		27.78	34.72	30.46
	Executive			28.95	36.11	43.06		38.89		36.75
	Board of Directors			31.30	33.33	41.67				35.43
	Other	33.33		39.58	22.92		38.89			33.68

Average Correct		Black (M)	Black (F)	White (M)	White (F)	Asian (M)	Asian (F)	Mixed (M)	Mixed (F)	Average Correct by Education Level
Education Level	GCSE/High School			29.52	27.31					28.42
	A/O Levels			30.34	28.59	33.33		25	33.33	30.12
	Degree	33.33	25	34.14	31.23	31.75	38.89	30.56	48.61	34.19
	Post Graduate Degree	28.70		31.79	32.91		47.22	38.89		35.90
	Other			31.17	34.48	47.22				37.62

Average Correct		Black (M)	Black (F)	White (M)	White (F)	Asian (M)	Asian (F)	Mixed (M)	Mixed (F)	Average Correct by Generagtion (Age)
Generation (Age)	Gen Z	34.92		34.04	33.3	25	47.22	36.11		35.09833333
	Millenials	30.55	25	35.24	31.28	38.88		25	61.11	35.29428571
	Gen X	30.55		31.62	30.35	29.62	38.88		34.72	32.62333333
	Baby Boomers			28.64	25.46	47.22		38.88		35.05

Correct Answers		White	Black	Asian	Mixed
Gender	Male	32.32	32.07	33.64	31.02
	Woman	31.21	25	44.44	43.52
	Prefer not to say	32.36			

Figure 4.6: *Heatmap showing average accuracy at detecting deepfakes by category.*

Interviews

This section sets out the findings and themes found through thematic analysis of the Interviews conducted on Cybersecurity professionals. Following Braun and Clarke's Six-Phase Model, five themes appeared in line with the research questions (Table 4.5).

Research Question	Theme
RQ1: How aware are employees and executives of deepfake threats?	**Theme 1:** Public Perceptions
RQ3: What are the primary risks deepfakes pose to corporations?	**Theme 4:** Risks Associated with Deepfakes for Companies
RQ4: How prepared are companies to detect and respond to deepfake-related threats, and what strategies are most effective?	**Theme 2:** Current and Future Threats of Deepfakes **Theme 3:** Corporate Preparedness and Response Measures **Theme 5:** Measures companies have put in place

Table 4.5: *Research Questions 1, 3 and 4*

Theme 1: *Public Perceptions of Deepfakes*

The findings indicate that while awareness of deepfakes exists among employees and executives, this awareness is often superficial and primarily associated with social media or entertainment, rather than corporate security risk (Table 4.6).

"Everyone knows the word deepfake, and most people know it in the context of Taylor Swift of Brad Pitt or Tom Cruise, which trivialises the threat."	Interviewee 5, CEO at a Deepfake Fraud Prevention Firm and Public Speaker

"They'll think from a social point of view that deepfakes are funning because they've seen face swap apps."	Interviewee 3, Chief Technologist with expertise in networking and security.
"I think everyone is aware, but they feel like they don't believe it can happen until it happens to them."	Interviewee 1, Pre-Sales Lead in Security, bridging technical expertise with business needs.

Table 4.6: *Deepfakes as entertainment*

Alternatively, Interviewee 9, a deepfake artist, mentioned the positives of deepfakes for entertainment (Table 4.7).

"Deepfakes are a benefit to the entertainment industry. This is just one example. Another thing is to relive or resurrect long gone actors." *"That was the work I am most proud of, bringing Elvis back to life digitally."*	Interviewee 9, 3D Artist specialising in state-of-the-art deepfake techniques

Table 4.7: *Positives of Deepfakes*

Executives and cybersecurity professionals usually have a higher level of awareness, whereas general employees are less informed about the risks deepfakes pose in corporate environments (Table 4.8).

"I don't know any CISO of a large organisation who won't be aware of the implications of deepfakes… But the broader employee base, less so."	Interviewee 3, Chief Technologist with expertise in networking and security.

"There is a general awareness that they are high risk simply because of what we are seeing in the media, with social media tools. But when it comes to corporate security, many still do not grasp the risk."	Interviewee 4, Chief Information Security Officer with expertise in cybersecurity.

Table 4.8: *Executives are more aware of the risk*

Despite growing awareness, employees often overestimate their ability to detect deepfakes. Employees do not understand how big the threat is (Table 4.9):

"We showed them ten portraits, and we asked them to evaluate which ones are AI-generated and which ones are not. Most of the people got less than five."	Interviewee 6, Co-Founder at Deepfake Detection Company
"I think a lot of them are aware that deepfakes are a threat, right? I don't think they are very aware on how to address it. And I don't think people know what they look and feel is like, right?"	Interviewee 8, Founder at a Deepfake Detection and Awareness Company
"The idea that someone at some point in your organisation is going to take the path of least resistance for their own convenience, that is the idea of rational choice theory."	Interviewee 1, Pre-Sales Lead in Security, bridging technical expertise with business needs.

Table 4.9: *Employees overestimate the threat*

Furthermore, larger organisations with dedicated security teams and bigger budgets tend to have more knowledge surrounding deepfakes, while smaller organisations often lack formal training on deepfake threats (Table 4.10).

"A lot of organisations, especially smaller ones, won't have any idea. They might have seen something on the news but will not have drawn the parallels." *"Smaller organisations don't have the time, resources or money to implement these things."*	Interviewee 2, Cyber Threat Hunter with expertise in Digital Forensics and Malware Analysis.

Table 4.10: *Company size affects preparedness*

Finally, deepfakes exploit human trust in visual and audio information, making individuals inherently vulnerable (Table 4.11).

"We are wired to trust authority… deepfakes turn that on its head, saying 'Do not trust everything digitally.' "	Interviewee 5, CEO at a Deepfake Fraud Prevention Company and Public Speaker.
"If I can't trust the person Infront of me is real, then I start to lack confidence in the whole fabric of humanity"	Interviewee 3, Chief Technologist with expertise in networking and security.

Table 4.11: *Human exploitation*

Theme 2: *Current and Future Threats of Deepfakes*

One of the biggest concerns of deepfakes is fraud, where deepfake technology is used to impersonate senior executives and authorise fraudulent transactions (Table 4.12).

"We are seeing Teams calls with a deepfake image of a senior employee asking for a payment to be made."	Interviewee 2, Cyber Threat Hunter with expertise in Digital Forensics and Malware Analysis.
"I think the number one risk is financial fraud... pretending to be from a supplier who wants to be paid or from a customer requesting an order."	Interviewee 4, Chief Information Security Officer with expertise in cybersecurity.

Table 4.12: *The risk of Deepfake Fraud*

The rise in remote working has further increased this risk, as employees may have never met executives in person and are therefore more susceptible to impersonation attacks (Table 4.13).

"With more people working from home, it is harder to get in-person verification."	Interviewee 2, Cyber Threat Hunter with expertise in Digital Forensics and Malware Analysis.
"Employees that trust audio and video streams will now need additional verification layers to establish trust."	Interviewee 7, Co-Founder at a deepfake detection company

Table 4.13: *Risk of remote working*

A key concern raised was how accessible and sophisticated deepfake technology is becoming, making it easier for attackers to launch highly realistic scams (Table 4.14).

"For $20 a month, you can get very good tools that allow you to create some of these artifacts." *"There is no debate. It is the largest growing threat vector."*	Interviewee 5, CEO at a Deepfake Fraud Prevention Company and Public Speaker.
"I can now download DeepSeek and run it on my laptop… I can create a deepfake attack way better than before and do it in seconds."	Interviewee 3, Chief Technologist with expertise in networking and security.
"It is simply good if I get my hands on the required footage, I can pretty much do whatever." *"Every improvement makes it more seamless and harder to detect."*	Interviewee 9, 3D Artist specialising in state-of-the-art deepfake techniques

Table 4.14: *Accessibility and Sophistication of Deepfakes*

Theme 3: *Corporate Preparedness and Response Measures*

Most interviewees acknowledged that corporate preparedness is extremely low (Table 4.15).

"Not even close to being ready for this, simply because we do not understand it". *"Less than 10% of companies would have even updated their policies to consider AI as a threat, and less than 1%, if not zero, will have anything about how they can*	Interviewee 4, Chief Information Security Officer with expertise in cybersecurity.

combat deepfakes apart from an initial reactive approach'.	
"Not at all. Not at all. I think most organisations are not prepared"	Interviewee 8, Founder at a Deepfake Detection and Awareness Company.
"Companies are not yet adopting solutions"	Interviewee 7, Co-Founder at a deepfake detection company.

Table 4.15: *Companies are unprepared*

Existing detection technology is still limited, with few tools able to detect deepfakes with the rapidly evolving landscape (Table 4.16).

"There's not a mass of tools I can buy right now that says this software has a 95% efficacy rate of detecting deepfakes."	Interviewee 3, Chief Technologist with expertise in networking and security.
"The technology used to launch an attack is much more powerful than the technology used to detect or prevent the attack."	Interviewee 5, CEO at a Deepfake Fraud Prevention Company and Public Speaker.
"The deepfake creators and the deepfake detection tools will constantly evolve against each other."	Interviewee 4, Chief Information Security Officer with expertise in cybersecurity.

Table 4.16: *Limited detection frameworks*

Due to challenges in digital verification, some companies are shifting to psychical verification methods such as multi-factor authentication and in-person verification (Table 4.17).

"For certain transactions, you'll have to go into a bank branch to verify in person, it is the safest way."	Interviewee 3, Chief Technologist with expertise in networking and security.
"The action to get employees to do is to develop a mental blocker, validating that this person is who they say they are."	Interviewee 1, Pre-Sales Lead in Security, bridging technical expertise with business needs.

Table 4.17: *Digital verification is untrustworthy*

Despite increasing awareness of deepfakes, employee awareness and training programmes remain scarce (Table 4.18).

"There is not a proper training module yet, but there has been awareness through emails and SharePoint Pages."	Interviewee 2, Cyber Threat Hunter with expertise in Digital Forensics and Malware Analysis.

Table 4.18: *Limited training and awareness programmes*

Theme 4: *Risks Associated with Deepfakes for Companies*

The financial risk posed by deepfakes was repeatedly emphasised, with deepfake-enabled fraud considered a high-reward strategy for attackers (Table 4.19).

"A deepfake attack, if successful, is five times more profitable than a standard email compromise scam." *"17% of employees will actually do something that can lead to ransomware, breach, or significant financial loss."*	Interviewee 8, Founder at a Deepfake Detection and Awareness Company.

"If we send money to a fraudulent account due to a deepfake, it could take 90 days before anyone realises." *"Your share price goes down and X can't pay off his mortgage. No one wants that."*	Interviewee 2, Cyber Threat Hunter with expertise in Digital Forensics and Malware Analysis.

Table 4.19: *High financial risk*

Furthermore, the potential for reputational damage was another key concern (Table 4.20).

"If a deepfake does the opposite of your brand vales say a trusted brand like Lego, it can cause lasting damage."	" Interviewee 3, Chief Technologist with expertise in networking and security.
"You could have someone pretending to be a part of the company when they are not, representing our views in the same was as fake Donal Trump or Fake presidents."	Interviewee 4, Chief Information Security Officer with expertise in cybersecurity.

Table 4.20: *Risk of reputational damage*

Furthermore, deepfakes can erode trust in the company (Table 4.21).

"If we were very trusting before in Teams or Zoom chats, and now we can't trust them anymore, there might be an impact."	Interviewee 4, Chief Information Security Officer with expertise in cybersecurity.
"If you are getting battered with deepfake requests, the time your boss actually calls you and says, 'I need you to do this,' you are probably not going to trust it or believe it."	Interviewee 2, Cyber Threat Hunter with expertise in Digital Forensics and Malware Analysis.

Table 4.21: *Erosion of trust*

Employees were highlighted as being prime targets of social engineering attacks, although this depends on digital literacy (Table 4.22).

"Older employees may be more vulnerable to social engineering attacks due to their level of digital skills."	Interviewee 7, Co-Founder at a deepfake detection company.
"Attackers use deepfakes to trick helpdesks into resetting credentials by impersonating employees in distress."	Interviewee 3, Chief Technologist with expertise in networking and security.
"At the moment, the onus is really on the user. If your user does not spot this, you are done."	Interviewee 2, Cyber Threat Hunter with expertise in Digital Forensics and Malware Analysis.
"The uncanny valley that you get with deepfakes is enough to allow humans to judge whether it is right or not."	Interviewee 4, Chief Information Security Officer with expertise in cybersecurity.

Table 4.22: *Social Engineering Attacks*

The quickly evolving deepfake threat landscape was noted as a high risk (Table 4.23).

"What is secure today, will be broken tomorrow. And you will have to raise your defenced, for them to be broken again."	Interviewee 5, CEO at a Deepfake Fraud Prevention Company and Public Speaker.

Table 4.23: *Fast pace deepfake threat landscape*

Theme 5: *Measures Companies have put in place*

Findings indicate that few companies have implemented specialised deepfake detection tools, with most relying on existing cybersecurity frameworks rather than deepfake-specific defences (Table 4.24).

"Few companies specialise in deepfake prevention, and adoption is low."	Interviewee 1, Pre-Sales Lead in Security, bridging technical expertise with business needs.
"The biggest misconception I see is that companies are overconfident, right? They think that their procedures and processes will hold up."	Interviewee 8, Founder at a Deepfake Detection and Awareness Company.

Table 4.24: *Reliance on pre-existing frameworks*

Some companies are exploring AI tools, but adoption remains limited due to cost and uncertainty around effectiveness. Others are implementing training awareness (Table 4.25).

"Companies using tools like Prompt Security can tailor defences to their needs." "Most vendors just bolt AI onto existing security products rather than creating something truly deepfake-specific."	Interviewee 1, Pre-Sales Lead in Security, bridging technical expertise with business needs.
"We do workshops for C-level management where they try to identify real vs AI-	Interviewee 6, Co-Founder at Deepfake Detection Company.

generated images, and they realise how difficult it is."	
"We have implemented deepfake training as part of our next awareness campaign."	Interviewee 4, Chief Information Security Officer with expertise in cybersecurity.

Table 4.25: *Adoption is happening, but slowly*

The findings indicate that although deepfake awareness exists among employees and executives, it remains inconsistent. Employees often associate deepfakes with entertainment, rather than recognising them as corporate security threats. In contrast, executives and cybersecurity professionals demonstrate higher awareness, yet this knowledge is not always effectively communicated across organisations. This gap affects corporate security strategies, as many companies focus on traditional cybersecurity measures while underestimating the unique challenges that deepfakes pose.

Deepfakes present significant security, financial, and reputational risks to corporations. Financial fraud, particularly deepfake-enabled impersonation scams, is already occurring. Beyond financial threats, deepfakes erode trust in corporate communication, with individuals questioning the authenticity of digital interactions. The potential for reputational damage is substantial, as deepfakes could be used to misrepresent employees, executives and companies.

Despite risks, corporate preparedness remains limited. While some companies have begun exploring detection tools and training, deepfake-specific countermeasures are not yet widely adopted. Many organisations rely on employees' accuracy to detect deepfakes, placing pressure on human judgement and the risk of social engineering. Overconfidence in existing security controls further exacerbates the issue.

These findings highlight the need for structured strategies, including greater awareness, the integration of detection tools, and stronger verification processes. The implications of these findings are explored in the discussion chapter.

5. Discussion

While deepfakes will become beneficial for the entertainment industry, the threat by criminals outweighs the benefits. This section discusses the implications of the findings from both statistical and thematic analysis. This study aimed to assess the risks of deepfakes in corporate environments by combining a quantitative survey, measuring employees' detection abilities, with qualitative insights from cybersecurity professionals. A mixed methods approach was essential to fill the research gap by bridging numerical evidence with expert viewpoints (Cresswell, 2017; Creswell & Tashakkori, 2007; Waters, 2019). The findings indicate that employees' ability to detect deepfake images accurately remains low (M=32.40%, SD= 14.2), far below the benchmark (M=53.16%). Furthermore, the findings many companies remain unprepared. Public perception associates deepfakes with entertainment rather than corporate risk. Attacks use deepfakes for fraud, posing security and reputational threats. Some organisations are implementing deepfake detection tools, but responses are reactive rather than proactive. Employees falling for deepfakes face severe consequences, highlighting the need for preventive strategies.

Scholars highlight the opportunities deepfakes present for the entertainment sector (Figure 3.1). Olivia described feeling proud of generating a deepfake of Elvis Presley performing a German song (Raziel, 2021). However, Amelia noted that employees often see deepfakes as 'funny', which is supported by public engagement in AI-generated content (Appendix A). Nico added this 'trivialises the threat' as deepfakes are categorised for celebrities rather than companies, aligning with Schreiber and Schreiber's (2024) argument that employees lack awareness about deepfake risks. Additionally, Amber emphasises that 17% of employees take actions that lead to financial loss (Breacher.ai, 2025). Amelia added that executives should be more aware of deepfakes than employees, considering 25.9% of executives say their companies have experienced deepfake incidents (Deloitte, 2024). However, a Factorial ANOVA found no significant differences in accuracy

between job levels, or any other demographics. This is supported by Ahmed and Chua (20203), finding no correlation between accuracy and demographics. Furthermore, a one-sample t -test revealed mean accuracy (32.36%) was significantly lower than the hypothesised benchmark, with large effect size (Cohen's D = 1.81). However, some of these studies incorporated incentives (Diel, et al., 2024a), in which Rathje et al. (2023) suggest that financial incentives motivate judgements of misinformation. Thus, while participants underperformed, these results may more accurately reflect real-world detection abilities (Abdelazeem, et al., 2022). This challenges Munk's (2015) argument that awareness increases the ability to recognise and mitigate cyber risks. Therefore, further large-scale research is needed to test these assumptions and explore the real-world applicability of these findings, providing a more comprehensive basis for future arguments.

Attribution Theory (Heider, 1958; Jones, et al., 1972; Weiner, 1974; Weiner, 1986) explains how individuals assign responsibility, with delayed on hidden responses often worsening consequences (Shim & Yang, 2016; Sjovall & Talk, 2004). Amelia noted that when well-established brands contradict their values, the impact can be long-lasting. Vâlsan et al. (2022) emphasise that misinformation has immediate effects, reinforcing findings by Grier and Forehand (2003) and Walker (2005) on its role in eroding trust and fostering public scepticism. Clara noted that customer loyalty may decline if deepfakes associate a company with a divisive public figure with opposing views, impacting employees. Clara adds that a lack of trust in video calls can weaken executive influence, supported by Vince stating it disrupts internal communication, a concern echoed by the European Parliament (2021) regarding deepfake-related psychological harm. Nico suggests we are wired to trust authority, therefore companies must respond swiftly (Vecchietti, et al., 2025) and maintain transparency (Knight & Nurse, 2020), failure to do so allows misinformation to spread unchecked, further blurring the line between truth and deception (Geddes, 2020; Tahir, et al., 2021), resulting in a lack of employee confidence in the whole fabric of humanity, says Amelia. However, a paired-samples t-test found that

participants were significantly more accurate in identifying real images that deepfake images (M=4.41%), suggesting real images might have clearer distinguishing features. Scholars support this in suggesting deepfake risks are exaggerated (Acerbi, et al., 2022; Altay, et al., 2023; Ecker, et al., 2022; Mercier, 2020; Simon, et al., 2023).

Clara stated that organisations are not even close to being ready for deepfakes, simply because we do not understand it. This is supported by McLuhan's (1975) Technological Determinism Theory suggesting the development of deepfakes forces corporations to react, as the deepfake algorithms constantly evolve, Clara adds. This is supported by neuromorphic computing as deepfake systems mimic human cognition (Mehonic, et al., 2020; Mehonic & Kenyon, 2022), intensifying security concerns (Karras, et al., 2020; Perov, et al., 2020; Widder, et al., 2022), as the technology used to launch an attack is more powerful than the technology used to detect it, a concern echoed by Nico. There are currently a limited number of deepfake detection companies and legislation across the world (Table 3.9,3.10). Amelia adds there is a lack of software that has a 95% efficacy rate of detecting deepfakes, despite increasing accessibility and high-resolution deepfakes, further increasing risks (Karras, et al., 2020; Perov, et al., 2020; Widder, et al., 2022). Furthermore, Irené comments that there is low adoption of deepfake measures, as companies prioritise other cyber risks (Eling, et al., 2021; Entrust, 2024) Amber added that most organisations are not prepared and suggested that companies may be overconfident with their current cybersecurity. This supports Beck's (1992) Risk Society Theory, suggesting that societies underestimate emerging technological risks. However, Rubén performs workshops for C-level management, and Clara has deepfake training as part of an upcoming awareness campaign. Multiple researchers support this, suggesting that training programmes reduce risks, as human error remains the biggest risk (Aldawood & Skinner, 2019, p. 73; Ghafir, et al., 2016; Salahdine & Kaabouch, 2019). Additionally, past studies on human detection accuracy found that training and assistance improves accuracy (Boyd, et al., 2023; Cartella, et al., 2024; Diel, et al., 2024b; Hulzebosch, et al., 2020; Kramer & Cartledge, 2024; Robertson, et al., 2018). However, a Pearson's

correlation suggested that employees with high confidence were not necessarily more accurate. Vince argues that the onus is on the user, which is supported by Schultz (2005), arguing that security threats are inherent to human-managed systems. Irené emphasises that employees need to develop a mental blocker to validate the individual, as the phrase "seeing is believing" is no longer reliable (Geddes, 2020; Tahir, et al., 2021). However, despite mundane visual communication (Costa, et al., 2020), Clara suggests the uncanny valley that you get with deepfakes is enough to allow humans to judge whether it is right or not. However, the introduction of deep learning has made detection difficult (Goodfellow, et al., 2014; Goodfellow, et al., 2016; Goodfellow, et al., 2020). Amelia states that there is no debate that deepfakes are the largest growing threat vector, with deepfake-related crime surging by 3000% (Entrust, 2024; Karras, et al., 2020; Perov, et al., 2020; Widder, et al., 2022). The rise of Open-source AI tools has lowered the barrier to creating deepfakes (Github, 2024). Olivia mentioned that getting their hands on the required footage means being able to create pretty much whatever, with over 95% of deepfake videos created on DeepFaceLab, intensifying security concerns. Critics suggest the high costs of ultra-realistic deepfakes make them unlikely for corporate fraud (Kaspersky Daily, 2023; Kaspersky Threat Intelligence Portal, 2024). However, for $20 a month, you can get very good tools that allow you to create some of these artefacts. Furthermore, this is supported by Becker's (1968) Theory of Rational Choice as deepfakes-as-a-service optimise profitability and low risk, therefore increasing the likelihood of attacks (Europol, 2022), as Irené states "due to someone taking a path of least resistance".

Despite biometric security traditionally being a secure method of MFA, deepfakes can recreate images with a small amount of content (Bontcheva, et al., 2024; Kietzmann, et al., 2020; Microsoft, 2025a; Wang, et al., 2023). However, Bhargav-Spantzel et al. (2006) argue that multiple layers of verification are adequate, supported by India "they establish trust", although Vince mentions that sending money to fraudulent accounts could take 90 days before anyone realises. This raises concerns about the long-term sustainability of

biometric authentication (Karras, et al., 2020; Perov, et al., 2020; Widder, et al., 2022), as Clara mentions, financial fraud is prominent with criminals pretending to be suppliers or customers demanding transactions. Amelia believes eventually individuals may have to verify confidential information in person.

6. Conclusion and Recommendations

This conclusion follows Trzeciak and Mackay's (1994) guidelines for a conclusion. The research explored the risks deepfake technology pose in corporate environments by addressing four questions: RQ1: How aware are employees and exes of deepfake threats? RQ2: How effectively can individuals distinguish between real and deepfake images, and what factors influence detection accuracy? RQ3: what are the primary risks deepfakes pose to corporations? RQ4: How prepared are companies to detect and respond to deepfake-related threats, and what strategies are most effective? The findings suggest that deepfake technology is a risk that many companies remain unprepared to handle. Employees often overestimate their ability to detect deepfakes, and corporate policies lack formal measures to counteract the increasing sophistication of deepfakes.

Each chapter of this dissertation has contributed to building a comprehensive understanding of deepfake risks. The introduction established the context and significance of deepfake threats, outlining the study's objectives. The methodology chapter detailed the mixed-methods approach, combining quantitative data (employee surveys) and qualitative insights (cybersecurity professionals). The literature review explored the evolution of deepfakes, corporate vulnerabilities and existing cybersecurity responses, identifying a significant research gap concerning corporate preparedness. The findings chapter revealed employee deepfake detection averaged 32.4% accuracy, below the 53.16% benchmark. Additionally, the study revealed that deepfakes are being exploited for financial fraud, reputational damage and social engineering, yet corporate responses remain largely reactive. This study has real world benefits in the corporate environment and an indication that the issue has become systemic and an epidemic. Therefore, adding a valuable gap in the research and providing a justification for the completion of a larger scale study. The discussion chapter critically analysed these findings in relation to existing research, emphasising the disconnect between perceived and actual deepfake detection abilities, alongside the psychological implications of deepfakes.

A key deduction is that corporate vulnerabilities to deepfake attacks is increasing more rapidly than mitigation strategies. Despite rising concerns over AI-driven deception, many businesses remain unprepared. This research highlights an urgent need for companies to adopt AI-driven detection tools, integrate awareness training into cybersecurity protocols, and establish clear response policies. Without such measures, organisations risk falling victim to sophisticated fraud, misinformation and reputational damage.

To address these vulnerabilities, this dissertation proposes the D.E.T.E.C.T Framework, a structured model integrating technology, policy and training into cybersecurity strategies. Companies should deploy AI-powered tools capable of analysing pixel inconsistencies and facial distortions. Evaluation processes should be strengthened through robust verification methods, ensuing that sensitive communications are secure. Employee training should be mandatory, incorporating interactive deepfake recognition workshops to improve awareness. Additionally, employees much enforce clear deepfake policies, embedding them within existing protocols and crisis response strategies. A well-defined reporting system should be implemented, allowing employees to flag suspicious content. Finally, businesses must track and adapt to evolving deepfake threats to ensure a proactive response. By integrating D.E.T.E.C.T, organisations can move from reactive security measures to a proactive defence strategy against deepfakes.

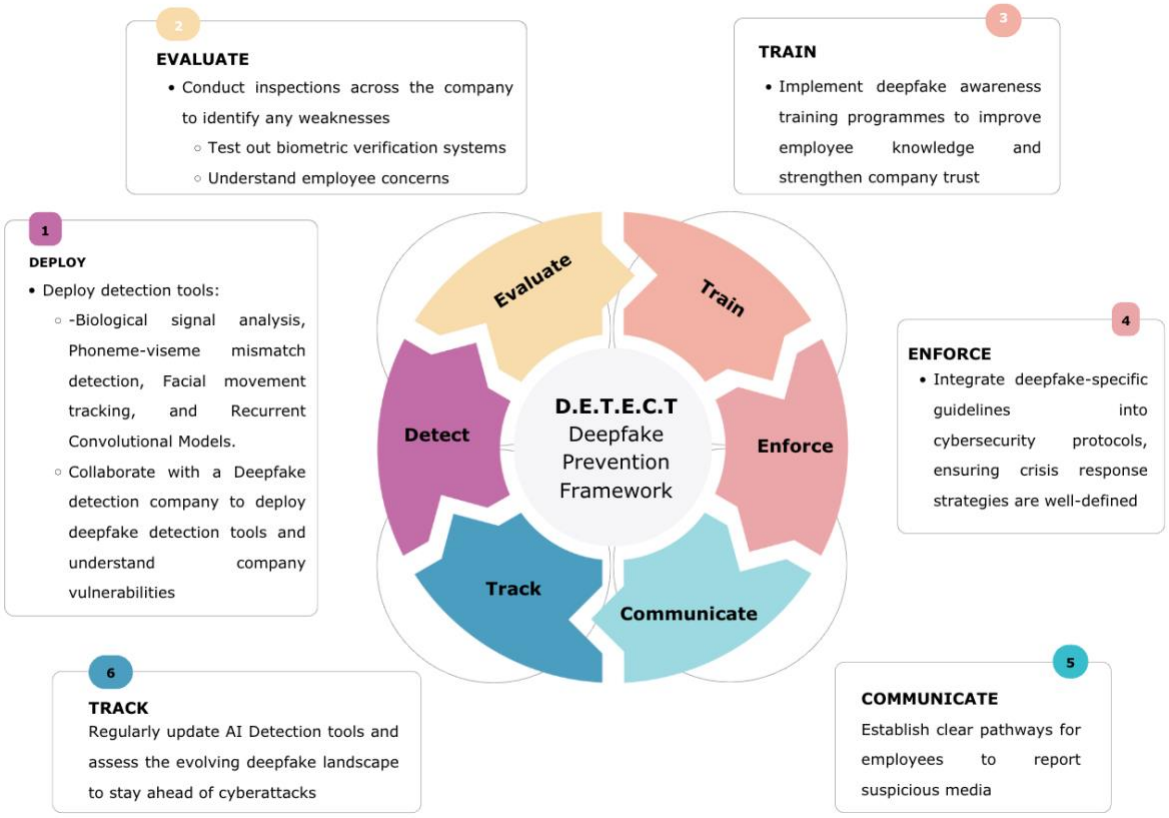


Figure 6.1: *Authors Own, D.E.T.E.C.T Deepfake Prevention Framework*

There are limitations to this research. The study relied on self-reported accuracy rates and subjective confidence measures, which may not fully capture real-world deepfake detection abilities. The sample for cybersecurity professionals was limited to nine participants, offering expert insights but not necessarily representing industry-wide perspectives. The study focused on static image-based deepfakes, excluding video and audio deepfakes which pose distinct challenges. Furthermore, participants prior knowledge of deepfakes was not assessed, nor was a control group included. Future studies should incorporate video and audio formats, evaluate and assess training methods, assessing accuracy over time.

This dissertation underscores the urgent need for businesses to recognise deepfakes as a legitimate threat. Without effective tools, policies and employee training, organisations remain highly vulnerable to deepfake technology. The rapid advancement demands a

proactive response, where companies, cybersecurity professionals and policymakers work together to safeguard digital communications. Future research and innovation in AI security will be critical in ensuring that companies remain resilient against the evolving landscape of deepfakes.

Bibliography

Abdelazeem, B. et al., 2022. The effectiveness of incentives for research participation: A systematic review and meta-analysis of randomized controlled trials. *PloS one,* 17(4), p. e0267534.

Acerbi, A., Altay, S. & Mercier, H., 2022. Research note: Fighting misinformation or fighting for information. *Harvard Kennedy School (HSK) Misinformation Review,* 3(1).

ADL, 2023. *The Dangers of Manipulated Media and Video: Deepfakes and More*. [Online] Available at: https://www.adl.org/resources/article/dangers-manipulated-media-and-video-deepfakes-and-more

[Accessed 27 February 2025].

Agarwal, S. & Farid, H., 2021. Detecting Deep-Fake Videos from Aural and Oral Dynamics. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. s.l.:s.n., pp. 981-989.

Agarwal, S., Farid, H., Fried, O. & Agrawala, M., 2020. Detecting deep-fake videos from phoneme-viseme mismatches. *In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops,* pp. 660-661.

Agarwal, S. et al., 2019. Protecting World Leaders Against Deep Fakes. *In CVPR workshops,* Volume 1, p. 38.

Ahmad, S., Mahmud, M. & Bajwa, F. A., 2023. The impact of cyberattacks awareness on customers' trust and commitment: an empirical evidence from the Pakistani banking sector. *Information and Computer Security,* 31(5), pp. 635-654.

Ahmed, S. & Chua, H. W., 2023. Perception and deception: Exploring individual responses to deepfakes across different modalities. *Heliyon,* 9(10).

AI Video, 2025. *aivideoapp*. [Online] Available at: https://www.tiktok.com/@aivideoapp?_t=ZN-8uloKfU2I0U&_r=1 [Accessed 10 March 2025].

Alahmari, A. & Duncan, B., 2020. Cybersecurity Risk Management in Small and Medium-Sized Enterprises: A Systematic Review of Recent Evidence. *In 2020 international conference on cyber situatioal awareness, data analytics and assessment (CyberSA),* June.pp. 1-5.

Aldawood, H. & Skinner, G., 2019, p. 73. Reviewing cyber security social engineering training and awareness programs-Pitfalls and ongoing issues. *Future internet,* 11(3), p. 73.

Alharahsheh, H. H., 2020. A review of key paradigms: Positivism VS interpretivism. *Global Academic Journal of Humanities and Social Sciences,* 2(3), pp. 39-43.

Al-Khazraji, S. H., Saleh, H. H., Khalid, A. I. & Mishkhal, I. A., 2023. Impact of deepfake technology on social media: Detection, misinformation, and societal implications. *The Eurasia Proceedings of Science Technology Engineering and Mathematics,* Volume 23, pp. 429-441.

Almalki, S., 2016. Integrating Quantitative and Qualitative Data in Mixed Methods Research--Challenges and Benefits. *Journal of education and learning,* 5(3), pp. 288-296.

Almutairi, Z. & Elgibreen, H., 2022. A review of modern audio deepfake detection methods: challenges and future directions. *Algorithms,* 15(5), p. 155.

Altay, S. et al., 2023. A survey of expert views on misinformation: Definitions, determinants, solutions, and future of the field. *Harvard Kennedy School (HKS) Misinformation Review,* 4(4).

Alvarez, 2024. *Portrait of beautiful young woman with blond hair on white background stock photo*. [Online] Available at: https://www.istockphoto.com/photo/portrait-of-beautiful-young-woman-with-blond-hair-on-white-background-gm1934795401-556204234?searchscope=image%2Cfilm [Accessed 20 November 2024].

Andrade, C., 2020. The Limitations of Online Surveys. *Indian journal of psychological medicine,* 42(6), pp. 575-576.

Antoniou, A., 2019. *Zao's deepfake face-swapping app shows uploading your photos is riskier than ever*. [Online] Available at: https://repository.essex.ac.uk/25392/1/a.antoniou_conversation_images_online.pdf

Apex Heroes, 2025. *apexheroes.ai*. [Online] Available at: https://www.tiktok.com/@apexheroes.ai?_t=ZN-8uloJb7GHyw&_r=1 [Accessed 10 March 2025].

Bachman, R. D., Schutt, R. K. & Plass, P. S., 2021. *Fundamentals of research in criminology and criminal justice*. 5th Editions ed. Los Angeles: SAGE.

Bateman, J., 2022. *Deepfakes and Synthetic Media in the Financial System: Assessing Threat Scenarios, Carnegie Endowment for International Peace*. s.l.:Carnegie Endowment for International Peace.

Battleitout, 2025. *battleitout4*. [Online] Available at: https://www.tiktok.com/@battleitout4?_t=ZN-8uloGPUTiZK&_r=1 [Accessed 10 March 2025].

BBC, 2019. *Facebook lets deepfake Zuckerberg video stay on Instagram*. [Online] Available at: https://www.bbc.co.uk/news/technology-48607673 [Accessed 21 February 2025].

Becker, G. S., 1968. Crime and punishment: An economic approach. *Journal of Political Economy,* Volume 76, pp. 169-217.

Beck, U., 1992. *Risk Society: Towards a New Modernity*. 1st Edition ed. Munich: SAGE .

Benaroch, M., 2018. Real options models for proactive uncertaintly-reducing migi. *InformationSystems Research,* 29(2), pp. 315-340.

Beyond The Screen, 2025. *Beyond The Screen*. [Online] Available at: https://www.tiktok.com/@beyond_thescreen?_t=ZN-8ulo8j54kBa&_r=1 [Accessed 10].

Bhargav-Spantzel, A., Squicciarini, A. & Bertino, E., 2006. Privacy preserving multi-factor authentication with biometrics. *In Proceedings of the second ACM workshop on Digital identity management,* pp. 63-72.

Bode, L., Lees, D. & Golding, D., 2021. The Digital Face and Deepfakes on Screen. *Convergence: The International Journal of Reserach into New Media Technologies,* 27(4), pp. 849-854.

Bojanc, R. & Jerman-Blažič, B., 2008. An economic modelling approach to information security risk management. *International Journal of Information Management,* 28(5), pp. 413-422.

Bondar, O., 2023. *Important LinkedIn Statistics Data & Trends*. [Online] Available at: https://www.linkedin.com/pulse/important-linkedin-statistics-data-trends-oleksii-bondar-pqlie/ [Accessed 7 March 2025].

Bontcheva, K. et al., 2024. *Generative AI and disinformation: recent advances, challenges, and opportunities,* s.l.: s.n.

Bookstaver, M., 2021. Secondary data analysis. *The encycopedia of research methods in criminology and criminal justice,* Volume 2, pp. 531-534.

Borrett, M., Carter, R. & Wespi, A., 2014. How is cyber threat evolving and what do organisations need to consider?. *Journal of business continuity and emergency planning,* 7(2), pp. 163-171.

Boyd, A., Tinsley, P., Bowyer, K. & Czajka, A., 2023. The Value of AI Guidance in Human Examination of Synthetically-Generated Faces. *In Proceedings of the AAAI Conference on Artificial Intelligence,* 37(5), pp. 5930-5938.

Braun, V. & Clarke, V., 2013. *Successful qualitaitve research: a practical guide for beginners*. s.l.:SAGE Publications.

Bray, S. D., Johnson, S. D. & Kleinberg, B., 2023. Testing human ability to detect 'deepfake' images of human faces. *Journal of Cybersecurity,* pp. 1-18.

Breacher.ai, 2025. *Deepfake Cybersecurity Solutions*. [Online] Available at: https://breacher.ai/ [Accessed 13 March 2025].

Breacher.ai, 2025. *Deepfake Cybersecurity Solutions*. [Online] Available at: https://breacher.ai/ [Accessed 13 March 2025].

Bregler, C., Covell, M. & Slaney, M., 2023. Video Rewrite: Driving Visual Speech with Audio. *In Seminal Graphics Papers: Pushing the Boundaries,* Volume 2, pp. 715-722.

Brenner, S. W., 2012. Cybercrime and the law: Challenges, issues, and outcomes. *UPNE*.

Brewer, R. et al., 2019. *Cybercrime prevention: Theory and applications*. s.l.:Springer Nature.

Bryman, A., 1984, p. 77. The debate about quantitative and qualitative research: a question of method or epistemology?. *British journal of Sociology,* pp. 75-92.

Bryman, A., 2016. *Bryman's Social Research Methods*. 5th Edition ed. Oxford: Oxford University Press.

Brynjolfsson, E. & McAfee, A., 2014. *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. s.l.:WW Norton & company.

Bucher, J., 2021. Inductive Reasoning. *The Encyclopedia of Research Methods in Criminology and Criminal Justice,* Volume 1, pp. 200-204.

Burkell, J. & Gosse, C., 2019. Nothing new here: Emphasizing the social and cultural context of deepfakes. *First Moday,* 24(12).

California Assembly Bill, 2019, n. 730, c. 493. *Elections: deceptive audio or visual media*. California: s.n.

Campbell, D. T., 1988. Methodology and epistemology for social sciences: Selected Papers. *University of Chicago Press*.

Cartella, G., Cuculo, V., Cornia, M. & Cucchiara, R., 2024. Unveiling the truth: Exploring human gaze patterns in fake images. *IEEE Signal Processing Letters*.

Carter, N., 2014. The use of triangulation in qualitative research. *Number5/September 2014,* 41(5), pp. 545-547.

Castillo Camancho, I. & Wang , K., 2021. A comprehensive review of deep-learning-based methods for image forensics. *Journal of Imaging,* 7(4), p. 69.

Cecillie_Arcurs, 2017. *There is nothing a positive attitude can't fix*. [Online] Available at: https://www.istockphoto.com/es/foto/no-hay-nada-que-una-actitud-

positiva-no-pueda-arreglar-gm638497186-114514179

[Accessed 25 November 2024].

CEOshrek, 2025. *Shrek*. [Online] Available at: https://www.tiktok.com/@ceoshrek?_t=ZN-8ulmMvdo0WM&_r=1 [Accessed 12 March 2025].

Cheng, Y., Jaekuk, L. & Jinzhe, Q., 2024. 8 Crisis Communication in the Age of AI: Navigating Opportunities, Challenges, and Future Horizons. *Media and crisis communication,* pp. 172-194.

Chesney, B. & Citron, D., 2019. Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review,* 107(6), pp. 1753-1820.

Chi, H., Maduakor, U., Alo, R. & Williams, E., 2020. Integrating deepfake detection into cybersecurity curriculum. *In Proceedings of the Future Technologies Conference (FTC),* Volume 1, pp. 588-598.

China's Provisions on the Administration of Deep Synthesis of Internet-based Information Services, 2023, Article 10. *Obligates providers to review and manage synthesised content*. China: s.n.

China's Provisions on the Administration of Deep Synthesis of Internet-based Information Services, 2023, Article 11. *Requires mechanisms for debunking false information*. China: s.n.

China's Provisions on the Administration of Deep Synthesis of Internet-based Information Services, 2023, Article 14. *Enforces strict data and biometric information protection*. China: s.n.

China's Provisions on the Administration of Deep Synthesis of Internet-based Information Services, 2023, Article 16. *Requires labelling of AI-generated or edited content*. China: s.n.

China's Provisions on the Administration of Deep Synthesis of Internet-based Information Services, 2023, Article 7. *Requires service providers to implement security and algorithm assessments*. China: s.n.

China's Provisions on the Administration of Deep Synthesis of Internet-based Information Services, 2023, Article 9. *Mandates real identity verification for users*. China: s.n.

Ciftci, U. A., Demir, I. & Yin, L., 2020. Fakecatcher: Detection of synthetic portrait videos using biological signals. *IEEE transactions on pattern analysis and machine intelligence*..

Clarity, 2025. *Clarity AI*. [Online] Available at: https://www.getclarity.ai/ [Accessed 7 March 2025].

Clark, T., Foster, L., Sloan, L. & Bryman, A., 2021. *Bryman's social research methods*. 6th Edition ed. Oxford: Oxford University Press.

Coleman, A., 2019. *'Deepfake' app causes fraud and privacy fears in China*. [Online] Available at: https://www.bbc.co.uk/news/technology-49570418 [Accessed 25 February 2025].

Colman, B., 2025. *AI Voice Fraud's Growing threat to Financial Security*. [Online] Available at: https://www.linkedin.com/pulse/185-million-lost-deepfake-scam-reality-defender-pfvte/ [Accessed 28 February 2025].

Costa, A., Bakker, J. & Plucinska, G., 2020. How and Why It works: The Principles and History behind Visual Communication. *Medical Writing,* Volume 29, pp. 16-21.

Cresswell, J. W., 2017. *Research design: Qualitative, quantitative, and mixed methods approaches*. 5th Edition ed. London: SAGE Publications.

Creswell, A. et al., 2018. Generative adversarial networks: An overview. *IEEE signal processing magazine,* 35(1), pp. 53-65.

Creswell, J. W. & Tashakkori, A., 2007. Editorial: Exploring the Nature of Research Questions in Mixed Methods Research. *Journal of Mixed Methods Research,* 1(3), pp. 207-211.

Cupchik, G., 2001. Constructivist realism: An ontology that encompasses positivism and constructivism approaches to the social scienct. *In Forum Qualitatie Sozialforschung/Forum: Qualitative Social Research,* 2(1).

CX Today, 2024. *Voice Cloning Tech Is Breaking Customer Authentication Systems*. [Online]

Available at: https://www.cxtoday.com/contact-center/voice-cloning-tech-is-breaking-customer-authentication-systems/

[Accessed 19 March 2025].

Daddy Shark, 2025. *daddy_shark00*. [Online]

Available at: https://www.tiktok.com/@daddy_shark00?_t=ZN-8uloBBqwTeY&_r=1

[Accessed 10 March 2025].

Damiani, J., 2019. *A Voice Deepfake Was Used To Scam A CEO Out Of $243,000*. [Online]

Available at: https://www.forbes.com/sites/jessedamiani/2019/09/03/a-voice-deepfake-was-used-to-scam-a-ceo-out-of-243000/

[Accessed 27 February 2025].

Das, S. & Nayak, T., 2013. Impact of cybercrime: Issues and Challenges. *International journal of engineering sciences and emerging technologies,* 6(2), pp. 142-153.

Data Protection Act, 2018. *Data Protection Act 2018 c.12*. [Online] Available at: https://www.legislation.gov.uk/ukpga/2018/12/contents

de Rancourt-Raymond, A. & Smaili, N., 2023. The unethical use of deepfakes. *Journal of Financial Crime,* 30(4), pp. 1066-1077.

de Rancourt-Raymond, A. & Smaili, N., 2023. The unethical use of deepfakes. *Journal of Financial Crime,* 30(4), pp. 1066-1077.

De Seta, G., 2021. Huanlian, or changing faces: Deepfakes on Chinese digital media platforms. *Convergence,* 27(4), pp. 935-953.

Deagreez, 2020. *Close-up portrait of his he nice attractive content brunet guy qualified tech IT geek expert programmer isolated over bright vivid shine vibrant teal green blue turquoise color background stock photo*. [Online] Available at: https://www.istockphoto.com/photo/close-up-portrait-of-his-he-nice-attractive-content-brunet-guy-qualified-tech-it-gm1202930572-345546780?searchscope=image%2Cfilm

[Accessed 20 November 2024].

Deepfakes Accountability Act, 2023, H.R. 5586. *DEEPFAKES Accountability Act*. United States of America: s.n.

Defamation Act , 2013. *Defamation Act 2013 c.26*. [Online] Available at: https://www.legislation.gov.uk/ukpga/2013/26/contents

Deloitte, 2024. *Generative AI and the fight for trust*. [Online] Available at: https://www2.deloitte.com/content/dam/Deloitte/us/Documents/Advisory/us-

generative-ai-and-the-fight-for-trust.pdf

[Accessed 10 March 2025].

Demir, I. & Ciftci, U. A., 2021. Where do deep fakes look? synthetic face detection via gaze tracking. *In ACM symposium on eye tracking research and applications,* pp. 1-11.

Denney, A. S. & Tewksbury, R., 2013. How to Write a Literature Review. *Journal of Criminal Justice Education,* 24(2), pp. 218-234.

Denscombe, M., 2021. *The good research guide: research methods for small-scale social research projects.* 7th Edition ed. London: Open University Press.

Denzin, N. K., 1978. *The research act: A theoretical introduction to sociological methods*. New York: McGraw-Hill.

Deranged AI, 2025. *Deranged AI*. [Online] Available at: https://www.tiktok.com/@derangedai?_t=ZN-8uloABE7IbG&_r=1 [Accessed 10 March 2025].

Diel, A., Bäuerle, A. & Teufel, M., 2024b. Inability to detect deepfakes: Deepfake detection training improves detection accuracy, but increases emotional distress and reduces self-efficacy. *But Increases Emotional Distress and Reduces Self-Efficacy*.

Diel, A. et al., 2024a. Human performance in detecting deepfakes: A systematic review and meta-analysis of 56 papers. *Computers in Human Behaviour Reports,* Volume 16, p. 100538.

Dixon, S. J., 2024. *LinkedIn: distribution of global audiences 2024, by age group*. [Online] Available at: https://www.statista.com/statistics/273505/global-linkedin-age-group/ [Accessed 8 March 2025].

Doffman, Z., 2019. *Chinese Deepfake App ZAO Goes Viral, Privacy Of Millions 'At Risk'*. [Online]

Available at: https://www.forbes.com/sites/zakdoffman/2019/09/02/chinese-best-ever-deepfake-app-zao-sparks-huge-faceapp-like-privacy-storm/

[Accessed 25 February 2025].

Doyle, L., Brady, A.-M. & Byrne, G., 2009. An overview of mixed methods research. *Journal of Research in Nursing,* 14(2), pp. 175-185.

DreamWeaver_AI, 2025. *dreamweaver.pl*. [Online]

Available at: https://www.tiktok.com/@dreamweaver.pl?_t=ZN-8uloMsjL0g2&_r=1

[Accessed 10 March 2025].

Dredge, S., 2025. *Sony Music reveals 75k takedowns of AI-generated deepfakes*. [Online]

Available at: https://musically.com/2025/03/10/sony-music-reveals-75k-takedowns-of-ai-generated-deepfakes/

[Accessed 19 March 2025].

Driscoll, D. L., Appiah-Yeboah, A., Salib, P. & Rupert, D. J., 2007. Merging qualitative and quantitative data in mixed methods research: How to and why not. *Ecological and Environmental Anthropology,* 3(1), pp. 19-28.

Dsouza, D. S., Hajjar, A. E. & Jahankhani, H., 2024. Deepfakes in Social Engineering Attacks. In: *Space Law Principles and Sustainable Measures*. Cham: Springer Nature Switzerland, pp. 153-183.

Eadicicco, L., 2019. *There's a fake video showing Mark Zuckerberg saying he's in control of 'billions of people's stolen data,' as Facebook grapples with doctored videos that spread misinformation*. [Online]

Available at: https://www.businessinsider.com/deepfake-video-mark-zuckerberg-instagram-2019-6

[Accessed 21 February 2025].

Ecker, U. K. H. et al., 2022. The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology,* 1(1), pp. 13-29.

El Mokadem, S. S., 2023. The Effect of Media Literacy on Misinformation and Deep Dake Video Detection. *Journal of Arab Media and Society,* pp. 53-78.

El Mokadem, S. S., 2023. The Effect of Media Literacy on Misinformation and Deep Fake VIdeo Detection. *Arab Media & Society,* p. 35.

Eling, M., McShane, M. & Nguyen, T., 2021. Cyber risk management: History and future research directions. *Risk Management and Insurance Review,* 42(1), pp. 93-125.

Entrust, 2024. *2025 Identity Fraud Report,* s.l.: Entrust and Onfido.

Etikan, I. & Bala, K., 2017. Sampling and sampling methods. *Biometrics and Biostatistics International Journal ,* 5(6), p. 00149.

Etikan, I., Musa, S. A. & Alkassim, R. S., 2016. Comparison of convenience sampling and purposive sampling. *American journal of theoretical and applied statistics,* 5(1), pp. 1-4.

European Parliament, 2021. *Tacking deepfakes in European policy*. [Online] Available at: https://www.europarl.europa.eu/RegData/etudes/STUD/2021/690039/EPRS_STU(2021)690039_EN.pdf

[Accessed 9 March 2025].

Europol, 2022. *Facing Reality? Law Enforcement and the Challenge of Deepfakes, an observatory report from the Europol Innovation Lab,* Luxembourg: Publications Office of the European Union.

Farid, H., 2022. Creating, uding, misuding, and detecting deep fakes. *Journal of Online Trustt and Safety,* 1(4).

feellife, 2013. *Portrait of an older asian wowan adult - Stock Photo*. [Online] Available at: https://www.istockphoto.com/es/foto/retrato-de-una-mujer-mayor-asi%C3%A1tica-adultos-gm187236102-27624658

[Accessed 25 November 2024].

Flick, T. & Morehouse, J., 2011. Chapter 7 - Attacking the utility companies. In: *Securing the Smart Grid: Next Generation Power Security Grid*. s.l.:Elsevier, pp. 109-142.

Foerderer, J. & Schuetz, S. W., 2022. Data Breach Announcements and Stock Market Reactions: A Matter of Timing?. *Management Science,* 68(10), pp. 2065-7791.

France-Presse, A., 2019. *Chinese deepfake app Zao sparks privacy row after going viral*. [Online]

Available at: https://www.theguardian.com/technology/2019/sep/02/chinese-face-swap-app-zao-triggers-privacy-fears-viral

[Accessed 25 February 2025].

Güera, D. & Delp, E. E., 2018. Deepfake video detection using recurrent neural networks. *In 2018 15th IEEE international conference on advanced video and signal based surveillance (AVSS),* pp. 1-6.

Gamage, D. et al., 2022. Are deepfakes concerning? analyzing conversations of deepfakes on reddit and exploring societal implications. *In proceedings of the 2022 CHI conference on human factors in computing systems,* pp. 1-19.

Geddes, K., 2020. Ocularcentrism and deepfakes: Should seeing be believing?. *Fordham Intellectual Property, Media & Entertainment Law Journal,* Volume 31, p. 1042.

Generated Photos, 2024. *Unique, worry-free model photos*. [Online] Available at: https://generated.photos/

[Accessed 10 November 2024].

GetReal Labs, 2025. *Deepfakes, real consequences*. [Online] Available at: https://www.getreallabs.com/ [Accessed 7 March 2025].

Ghafir, I., Prenosil, V., Alhejailan, A. & Hammoudeh, M., 2016. Social engineering attack strategies and defence approaches. *2016 IEEE 4th international conference on future internet of things and cloud (FiCloud),* Volume IEEE, pp. 145-149.

Ghilom, M. & Latifi, S., 2024. The Role of Machine Learning in Advanced Biometric Systems. *Electronics,* 13(13), p. 2667.

Github, 2024. *iperov/ DeepFaceLab*. [Online] Available at: https://github.com/iperov/DeepFaceLab [Accessed 11 March 2025].

Given, L., 2025. *Generative AI and deepfakes are fuelling health misinformation. Here's what to look out for so you don't get scammed*. [Online] Available at: https://theconversation.com/generative-ai-and-deepfakes-are-fuelling-health-misinformation-heres-what-to-look-out-for-so-you-dont-get-scammed-246149 [Accessed 19 March 2025].

Given, L. M., 2008. *The SAGE encyclopedia of qualitative research methods*. s.l.:Sage Publications.

Glass, G. V., 1976. Primary, secondary and meta-analysis of research. *Educational researcher,* 5(10), pp. 3-8.

Global Times, 2023. *Chinese netizens stay alert over AI scam*. [Online] Available at: https://www.globaltimes.cn/page/202305/1291150.shtml [Accessed 28 February 2025].

Goodfellow, I., Bengio, Y., Courville, A. & Bengio, Y., 2016. *Deep Learning*. Cambridge: MIT Press.

Goodfellow, I. et al., 2014. Generative adversarial nets. *Advances in neural information processing systems,* Volume 7.

Goodfellow, I. et al., 2020. Generative Adversarial Networks. *Communications of the ACM,* 63(11), pp. 139-144.

Goodman, L. A., 1961. Snowball Sampling. *The annals of mathematical statistics,* pp. 148-170.

Gordon, S. & Ma, Q., 2003. *Convergence of virus writers and hackers: Fact or fantasy*. Cupertine, CA: Symantec Security White Paper.

GOV.UK, 2019. *Deepfakes and Audiovisual Information,* s.l.: Centre for Data Ethics and Innovation.

GOV.UK, 2023. *A pro-innovation approach to AI regulation*. [Online] Available at: https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper#correction-slip

[Accessed 7 March 2025].

Gregory, J., 2025. *Are successful deepfkae scams more common that we realize*. [Online]

Available at: https://securityintelligence.com/articles/are-successful-deepfake-scams-more-common-than-we-realize/

[Accessed 6 March 2025].

Grier, S. S. & Forehand, M. R., 2003. When is Honesty The Best Policy? The Effect of Stated Company Intent on Consumer Skepticism. *Journal of consumer psychology,* 13(3), pp. 349-356.

Guba, E. G. & Lincoln, Y. S., 1994. Competing paradigms in qualitative research. *Handbook of qualitative research,* 2(163-194), p. 105.

Guess, A., Nagler, J. & Tucker, J., 2019. Less than you think: Prevalance and predictors of fake news dissemination on Facebook. *Science advances,* 5(1).

Gui, J. et al., 2021. A review on generative adversarial networks: Algorithms, theory, and applications. *IEEE transactions on knowledge and data engineering,* 35(4), pp. 3313-3332.

Gwebu, K. L., Wang, J. & Wang, L., 2018. The Role of Corporate Reputation and Crisis Response Strategies in Data Breach Management. *Journal of Management Information Systems,* 35(2), pp. 683-714.

Halcomb, E. J., 2018. Mixed methods research: The issue beyond combining methods. *Journal of Advanced Nursing,* 75(3), pp. 499-501.

Hanson, B., 2008. Wither qualitative/quantitative?: grounds for methodological convergence. *Qual Quant,* 42(1), pp. 97-111.

Hartono, W. et al., 2024. Challenges of criminal investigation cybercrime. *Awang Long Law Review,* 7(1), pp. 11-19.

Hatim's Shorts, 2025. *hatimsshorts*. [Online] Available at: https://www.tiktok.com/@hatimsshorts?_t=ZN-8uloDup7QUo&_r=1 [Accessed 10 March 2025].

Headshot, 2025. *AI-Generated Photos for 3D Human Creation*. [Online] Available at: https://www.reallusion.com/character-creator/headshot/ai-generated-face.html [Accessed 8 March 2025].

Heap, V. & Waters, J., 2019. *Mixed Methods in Criminology*. 1st Edition ed. London: Routledge.

Heider, F., 1958. *The Psychology of Interpersonal Relations*. New York: WIley.

Hermosilla, G. et al., 2021. Thermal face generation using stylegan. *IEEE Access,* Volume 9, pp. 80511-80523.

Holdsworth, J. & Scapicchio, M., 2024. *What is deep learning?*. [Online] Available at: https://www.ibm.com/topics/deep-learning [Accessed 6 February 2025].

Holt, T. & Bossler, A., 2015. *Cybercrime in progress: Theory and prevention of technology-enabled offenses*. 1st Edition ed. London: Routledge.

Honor, 2024. *Honor*. [Online] Available at: https://www.honor.com/global/news/honor-shanghai-mwc-2024/ [Accessed 7 March 2025].

Hox, J. J. & Boeije, H. R., 2005. Data collection, primary vs secondary. *Encyclopedia of social measurements,* 1(1), pp. 593-599.

Huang , Q. & Maracic, J., 2024. Consumer perception of Deepfake Technology in Marketing: AN abductive study on consumer attitude, trust and brand authenticity. Issue Dissertation.

Hulzebosch, N., Ibrahimi, S. & Worring, M., 2020. Detecting CNN-generated facial images in real-world scenarios. *In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops,* pp. 642-643.

Hussein, A., 2009. The use of trianguation in social sciences research: Can qualitative and quanitative methods be combined?. *Journal of comparative social work,* 4(1), pp. 106-117.

Hwang, Y., Ryu, J. Y. & Jeong, S. H., 2021. Effects of disinformation using deepfake: The protective effect of media literacy education. *Cyberpsychology, Behaviour, and Social Networking,* 24(3), pp. 188-193.

IASME, 2025. *Inclusive cyber security*. [Online] Available at: https://iasme.co.uk/ [Accessed 7 March 2025].

Ikenga, F. A. & Nwador, A. F., 2024. The intersection of artificial intelligence, deepfake, and the politics of international diplomacy. *Ianna J. Interdiscip. Stud,* 6(2), pp. 53-70.

Information Technology Act, 2000, s. 66 (C) . *Punishment for Indetity Theft*. India: s.n.

Information Technology Act, 2000, s. 66 (D). *Punishment for cheating by personation by using computer resource*. India: s.n.

Inninois AI Video Interivew Act, 2024. *The Artificcial Intelligence Video Interview Act*. Illinois: s.n.

Izusek, 2023. *Headshot of happy young businessman stock photo*. [Online] Available at: https://www.istockphoto.com/photo/headshot-of-happy-young-businessman-gm1542565697-525855178?searchscope=image%2Cfilm [Accessed 17 November 2024].

Jafar, M. T., Ababneh, M., Al-Zoube, M. & Elhassan, A., 2020. Forensics and analysis of deepfake videos. *in 2020 11th international conference on information and communication systems (ICICS),* April.pp. 053-058.

Jain, A. K. & Kumar, A., 2012. Biometric recognition: an overview. *Second generation biometrics: The ethical, legal and socail context,* pp. 49-79.

Johnson, T. P., 2014. Snowball sampling: introduction. *Wiley StatsRef: Statsistics Reference Online*.

Johnston, M. P., 2014. Secondary data analysis: A method of which the time has come. *Qualitative and quantitative methods in libraries,* 3(3), pp. 619-626.

Jones, E. E. et al., 1972. *Attribution: Perceiving the Causes of Behavior.* Morristown, NJ: General Learning Press.

Junjie, M. & Yingxin, M., 2022. The discussions of Positivism and Interpretivism. *Online Submission,* 4(1), pp. 10-14.

Kalaiarasu, S., Rahman, N. A. A. & Harun, K. S., 2024. Deepfake impact, security threats and potential preventions. *In AIP Conference Proceedings,* 2802(1).

Karinshak, E. & Jin, Y., 2023. AI-driven disinformation: a framework for organisational preparation and response. *Journal of Communication Management,* 27(4), pp. 539-562.

Karras, T., Laine, S. & Aila, T., 2019. A style-based generator architecture for generative adversarial networks. *In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition,* pp. 4401-4410.

Karras, T., Laine, S. & Aila, T., 2024. *Style GAN2 Images*. [Online] Available at: https://thispersondoesnotexist.com/ [Accessed 20 November 2024].

Karras, T. et al., 2020. Analyzing and improving the image quality of stylegan. *In proceedings of the IEEE/CVF conference on computer vision and pattern recognition,* pp. 8110-8119.

Kaspersky Daily, 2023. *How real is deepfake threat? Understanding the mechanics of the darknet deepfake industry.* [Online] Available at: https://www.kaspersky.co.uk/blog/deepfake-darknet-market/25961/ [Accessed 25 February 2025].

Kaspersky Threat Intelligence Portal, 2024. *Threat Intelligence search feature*. [Online] Available at: https://tip.kaspersky.com/Help/Doc_data/en-US/TISearch.htm [Accessed 25 February 2025].

Kerner, C. & Risse, M., 2020. Beyond Porn and Discreditation: Epistemic Promises and Perils of Deepfake Technology in Digital Lifeworlds. *Moral Philosophy and Politics*.

Khan, M. R. et al., 2023. Exploring Neurophysiological Responses to Cross-Cultural Deepfake Videos.. *International ,* pp. 41-45.

Kietzmann, J., Lee, L. W., McCarthy, I. P. & Kietzmann, T. C., 2020. Deepfakes: Trick or treat. *Buziness Horizons,* 63(2), pp. 135-146.

Kim, J.-J. & Hong, S.-P., 2011. A method of risk assessment for multi-facotr authentication. *Journal of Information Proceedings Systems,* 7(1), pp. 187-198.

Knight, R. & Nurse, J. R. C., 2020. A framework for effective corporate communication after cyber security incidents. *Computers and Security,* Volume 99, p. 102036.

KnowBe4, 2024. *How a North Korean Fake IT Worker Tried to Infiltrate Us*. [Online] Available at: https://blog.knowbe4.com/how-a-north-korean-fake-it-worker-tried-to-infiltrate-us [Accessed 27 February 2025].

Korarkar, U. & Sakarkar, G., 2023. Unmasking Deepfakes Advancements, Challenges, and Ethical Considderations. *In International Conference on Electrical and Electronics Engineering,* pp. 249-262.

Kramer, R. S. S. & Cartledge, C., 2024. Crowds Improve Human Detection of AI-Synthesised Faces. *Applied Cognitive Psychology,* 38(5), p. 4245.

Krea.AI, 2025. *Generate Images*. [Online] Available at: https://www.krea.ai/ [Accessed 7 March 2025].

Krombholz, K., Hobel, H., Huber, M. & Weippl, E., 2015. Advanced social engineering attacks. *Journal of Information Security and Applications,* Volume 22, pp. 113-122.

Krumsvik, R. J., Jones, L. Ø., Øfstegaard, M. & Eikeland, O. J., 2016. Upper secondary school teachers' digital competence: Analysed by demographic, personal and professional characteristics. *Nordic journal of digital literacy,* 11(3), pp. 143-164.

Kwon, J. & Johnson, M. E., 2014. Health-care security strategies for data protection and regulatory compliance. *Journal of Management Information Systems,* 30(2), pp. 41-66.

Kyle, P., 2025. *Ai Opportunities Action Plan*. [Online] Available at: https://www.gov.uk/government/publications/ai-opportunities-action-plan/ai-opportunities-action-plan [Accessed 7 March 2025].

Kyle, P., Department for Science, I. a. T., Innovation, U. R. a. & Institute, A. S., 2024. *Research programme to ensure UK economy uses AI to grow safely*. [Online] Available at: https://www.gov.uk/government/news/research-programme-to-ensure-uk-economy-uses-ai-to-grow-safely [Accessed 7 March 2025].

Lakens, D., 2022. Sample size justification. *Collabra: psychology,* 8(1), p. 3326.

Lawrence, T., 2024. *The Rise of Deepfakes*. [Online] Available at: https://www.engagewithus.com/the-rise-of-deepfakes/ [Accessed 25 February 2025].

Lee, C.-J. G., 2012. Reconsidering constructivism in qualitative research. *Educational Philosophy and theory,* 44(4), pp. 403-412.

leezsnow, 2005. *Beautiful Afroamerican woman - Stock Photo*. [Online] Available at: https://www.istockphoto.com/es/foto/hermosa-mujer-afroamericana-gm182146354-1141650

[Accessed 25 November 2024].

Lemon Photo, 2024. *Waist up shot of a handsome hispanic Latino carefree black male looking at the camera with big smile stock photo*. [Online] Available at: https://www.istockphoto.com/photo/waist-up-shot-of-a-handsome-hispanic-latino-carefree-black-male-looking-at-the-gm2135643049-568240866

[Accessed 19 November 2024].

Leng, C. & Ho-him, C., 2024. *Arup lost $25mn in Hong Kong deepfake video conference scam*. [Online]

Available at: https://www.ft.com/content/b977e8d4-664c-4ae4-8a8e-eb93bdf785ea

[Accessed 6 March 2025].

leungchopan, 2016. *Portrait of a businessman - Stock Photo*. [Online] Available at: https://www.istockphoto.com/es/foto/retrato-de-un-hombre-de-negocios-gm608003812-104307365

[Accessed 23 November 2024].

Liao, X. et al., 2023. FAMM: facial muscle motions for detecting compressed deepfake videos over social networks. *IEEE Transactions on Circuits and Systems for Video Technology,* 33(12), pp. 7236-7251.

Lim, S. Y., Chae, D. K. & Lee, S. C., 2022. Detecting deepfake voice using explainable deep learning techniques. *Applied Sciences,* 12(8), p. 3926.

LinkedIn, 2025. *LinkedIn*. [Online] Available at: www.linkedin.com [Accessed 30 January 2025].

Liporazzi, B., 2025. *LinkedIn*. [Online] Available at: https://www.linkedin.com/in/bettina-liporazzi/ [Accessed 10 March 2025].

Liu, Z., Qi, X. & Torr, P. H., 2020. Global texture enhacement for fake face detection in the wild. *In Porceedings of the IEEE.CVF conference on computer vision and pattern recognition,* pp. 8060-8069.

Loti, 2025. *Online Protection for public figures*. [Online] Available at: https://goloti.com/ [Accessed 7 March 2025].

Lydon, L., 2021. *Corproate under reporting of cybercrime: Why does reporting to authorities matter?*. [Online] Available at: https://www.royalholloway.ac.uk/research-and-education/departments-and-schools/information-security/research/explore-our-research/isg-technical-reports/

Lyu, S., 2020. Deepfake detection: Current challenges and next steps. *In 2020 IEEE international conference on multimedia & expo workshops (ICMEW),* pp. 1-6.

Machi, L. A. & McEvoy, B. T., 2022. *The literature review six steps to success.* 4th Edition ed. Corwin: Thousand Oaks.

Magann, 2019. *Young bond woman stock photo*. [Online] Available at: https://www.istockphoto.com/photo/young-blond-woman-

gm1151933799-312333172?searchscope=image%2Cfilm

[Accessed 19 November 2024].

Magramo, K., 2024. *British engineering giant Arup revealed as $25 million deepfake scam victim*. [Online]

Available at: https://edition.cnn.com/2024/05/16/tech/arup-deepfake-scam-loss-hong-kong-intl-hnk/index.html

[Accessed 25 Febrary 2025].

Mandelcorn, S., Modarres, M. & Mosleh, A., 2013. *An Explanatory Model of Cyber-Attacks Drawn from Rational Choice Theory,* College Park: University of Maryland.

Manky, D., 2013. Cybercrime as a service: a very modern business. *Computer Fraud & Security,* Issue 6, pp. 9-13.

Marciano, J., 2024. *Evaluating SOC-2 Control CC1.4/COSO with the rise of Adversarial AI*. [Online]

Available at: https://ssrn.com/abstract=4940938 or http://dx.doi.org/10.2139/ssrn.4940938

[Accessed 27 February 2025].

Marsh, E. J., Cantor, A. D. & Brashier, N. M., 2016. Believing that Humans Swallow Spiders in Their Sleep: False Beliefs as Side Effects of the Processes that Support Accurate Knowledge. *Psychology of Learning and Motivation,* Volume 64, pp. 93-132.

Maruna, S., 2010. Mixed method research in criminology: Why not go both ways?. *Handbook of quantiative criminology,* pp. 123-140.

Masood, M. et al., 2023. Deepfakes generation and detection: State of the art, open challenges, countermeasures, and way forward. *Applied intelligence,* 53(4), pp. 3974-4026.

McLuhan, M., 1975. *Understanding the Media: Extensions of Man*. London and New York: s.n.

MEGA, 2025. *megakivideos*. [Online] Available at: https://www.tiktok.com/@megakivideos?_t=ZN-8uloCHNldRO&_r=1 [Accessed 10 March 2025].

Mehonic, A. & Kenyon, A. J., 2022. Brain Inspired computing needs a master plan. *Nature,* 604(7905), pp. 255-260.

Mehonic, A. et al., 2020. Memristors - From In-Memory Computing, Deep Learning Acceleration, and Spiking Neural Networks to the Future of Neuromorphic and Bio-Inspired Computing. *Advanced Intelligent Systems,* 2(11), p. 2000085.

Menon, V. & Muraleedharan, A., 2020. Internet-based surveys: relevance, methodological considerations and troubleshooting strategies. *General Psychiatry,* 33(5).

Mercier, H., 2020. *Not born yesterday: The science of who we trust and what we believe*. s.l.:Princeton University Press.

Microsoft, 2025a. *VALL-E: A neural codec language model for speech synthesis*. [Online] Available at: https://www.microsoft.com/en-us/research/project/vall-e-x/ [Accessed 27 February 2025].

Microsoft, 2025b. *What is: Multifactor Authentication*. [Online] Available at: https://support.microsoft.com/en-gb/topic/what-is-multifactor-authentication-e5e39437-121c-be60-d123-eda06bddf661

[Accessed 27 February 2025].

Miller, J. S., Birch, M., Mauthner, M. & Jessop, J., 2012. *Ethics in qualitative research*. s.l.:SAGE Publications.

Miller, M., 2021. *Deepfakes: Real threat,* s.l.: KPMG, Reality Defender.

Millett, L. I. & Pato, J. N., 2010. *Biometric recognition: Challenges and opportunities*. s.l.:s.n.

Milmo, D., 2024. *UK engineering firm Arup falls victim to £20m deepfake scam*. [Online] Available at: https://www.theguardian.com/technology/article/2024/may/17/uk-engineering-arup-deepfake-scam-hong-kong-ai-video
[Accessed 25 February 2025].

Ministry of Justice & Davis-Jones MP, A., 2025. *Government crackdown on explicit deepfakes*. [Online] Available at: https://www.gov.uk/government/news/government-crackdown-on-explicit-deepfakes
[Accessed 6 March 2025].

Mississippi Senate Bill, 2024, sb. 2577. *o Create Criminal Penalties For The Wrongful Dissemination Of Digitizations; And For Related Purposes*.. Mississippi: s.n.

Mitnick, K. & Simon, M. L., 2003. The art of deception. In: *Controlling the human element of security*. s.l.:John WIley & Sons.

Moczadlo, D., 2025. *LinkedIn*. [Online] Available at: https://www.linkedin.com/feed/update/urn:li:activity:7305307387110780928/
[Accessed 10 March 2025].

moneybusinessimages, 2015. *Mixed race businesswoman, head and shoulders portrait - Stock Photo*. [Online] Available at: https://www.istockphoto.com/es/foto/empresaria-de-raza-mixta-retrato-

de-cabeza-y-hombros-gm487807120-73545361

[Accessed 24 November 2024].

moneybusinessimages, 2024. *Head And Shoulders Portrait Of Smiling Man Outside Relaxing In Countryside Looking Off Camera stock photo*. [Online] Available at: https://www.istockphoto.com/photo/head-and-shoulders-portrait-of-smiling-man-outside-relaxing-in-countryside-looking-gm2150270850-571448050

[Accessed 19 November 2024].

Moreno, F. R., 2024. Generative AI and deepfakes: a human rights approach to tackling harmful content. *International Review of Law, Computers and Technology,* 38(3), pp. 297-326.

Mullen, M., 2022. A New Reality: Deepfake Technology and the World around US. *Mitchell Hamline Law Review,* Volume 1, pp. 210-234.

Munk, T., 2024. Concluding Remarks, the Merger of the Online and Offline Worlds. In: *Victimisation in the Digital Age*. London: Routledge, pp. 239-251.

Munk, T. H., 2015. Cyber-Security in the European Region: Anticipatory Governance and Practices. *The University of Manchester (United Kingdom)*.

Mustak, M. et al., 2023. Deepfakes: Deceptions, mitigations, and opportunities. *Journal of Business Research,* Volume 154, p. 113368.

Naffi, N. et al., 2025. Empowering youth to combat malicious deepfakes and disinformation: An experiential and reflective learning experience informed by personal construct theory. *Journal of Constructivist Psychology,* 38(1), pp. 119-140.

Nas, E. & de Kleijn, R., 2024. Conspiracy thinking and social media use are associated with the ability to detect deepfakes. *Telematics and Informatics,* Volume 87, p. 102093.

National Defense Authorisation Act (NDAA), 2021. *National Defense Authorisation Act for Fiscal Year*. United States of America: s.n.

National Stock Exchange of India, 2024. *Caution for Investors - fake videos of NSE MD and CEO Shri Ashishkuma Chauhan recommending stocks*. [Online] Available at: https://nsearchives.nseindia.com/web/sites/default/files/2024-07/PR_cc_10062024.pdf

[Accessed 14 March 2025].

NCSC, 2025. *The near-term imPct of AI on the cyber threat*. [Online] Available at: https://www.ncsc.gov.uk/pdfs/report/impact-of-ai-on-cyber-threat.pdf [Accessed 7 March 2025].

Ncubukezi, T., 2022. Human errors: A cybersecurity concern and the weakest link to small businesses. *Proceedings of the 17th Interntational COnferences on Information Warfare and Security,* p. 395.

Newman, I. & Benz, C. R., 1998. *Qualitative-quantitative research methodology: exploring the interactive continuum*. Carbondale, IL: Southern Illinois University Press.

Nichols, R., Rathgeb, C., Drozdowski, P. & Busch, C., 2022. Psychological Evaluation of Human Performance in Detecting Digital Face Image Manipulations. *IEEE Access,* Volume 10, pp. 31359-31376.

Nightingale, S. J. & Farid, H., 2022. AI-synthesized faces are indistinguishable from real faces and more trustworthy. *Proceedings of the National Academy of Sciences,* 119(8), p. 212048119.

Noto, G., 2024. *Scammers sophon $25M from engineering firm Arup via AI deepfake 'CFO'*. [Online] Available at: https://www.cfodive.com/news/scammers-siphon-25m-engineering-firm-

arup-deepfake-cfo-ai/716501/

[Accessed 25 February 2025].

O'Connor, C. & Weatherall, J. O., 2019. Why we trust lies. *Scientific American,* 321(3), pp. 54-61.

Ometov, A. et al., 2018. Multi-factor authentication: A survey. *Cryptography,* 2(1), p. 1.

O'Neil, C., 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. s.l.:Crown Publishing Group.

OpsAIDidItAgain, 2025. *OpsAIDidItAgain*. [Online] Available at: https://www.tiktok.com/@opsaididitagain?_t=ZN-8ulo7QW6qCE&_r=1 [Accessed 10 March 2025].

Oslen, W., 2004. Triangulation in social research: qualitative and quantitative methods can really be mixed. *Developments in sociology ,* Volume 20, pp. 103-118.

Packer, M. J. & Goicoechea, J., 2000. Sociocultural and Constructivist theories of learning: Ontology, not just epistemology. *Educational psychologist,* 35(4), pp. 227-241.

PacoRomero, 2012. *Asian businesswoman wearing traditional Chinese dress - Stock Photo*. [Online] Available at: https://www.istockphoto.com/es/foto/asi%C3%A1tica-empresaria-usando-vestido-chino-tradicional-gm155360699-19388560 [Accessed 25 November 2024].

Paravision AI, 2025. *Ethically developed AI building blocks for the next generation of identity*. [Online] Available at: https://www.paravision.ai/ [Accessed 7 March 2025].

Patton, M. Q., 1999. Enhancing the quality and credibility of qualitative analysis. *Health services research,* 34(5 Pt 2), p. 1189.

Peltier, T., 2006, p13. Social Engineering: Concepts and Solutions. *Information Security Journal,* 15(5), p. 13.

Perov, I. et al., 2020. DeepFaceLab: Integrated, Flexible and extensible face-swapping framework. *arXiv prepring arXiv:2005.05535*.

Petratos, P. N., 2021. Misinformation, disinformation, and fake news: Cyber risks to business. *Business Horizons,* 64(6), pp. 763-774.

Petratos, P. N., 2021. Misinformation, disinformation, and fake news: Cyber risks to businesses. *Business Horizons,* 64(6), pp. 763-774.

Photique, 2014. *Business portrait stock photo*. [Online] Available at: https://www.istockphoto.com/photo/business-portrait-gm495744619-41145126?searchscope=image%2Cfilm

[Accessed 14 November 2024].

Protecting Americans from Foreign Adversary Controlled Applications Act, 2024. *Protecting Americans from Foreign Adversary Controlled Applications Act*. United States of America: s.n.

Ragin, C. C., 1994. *Constructing social research: the unity and diversity of method*. Thousand Oaks, CA: Pine Forge.

Rana, S., Nobi, M. N., Murali, B. & Sung, A., 2022. *Deepfake Detection: A Systematic Literature Review*. [Online] Available at: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9721302

[Accessed 25 February 2025].

Randolph, J., 2019. A guide to writing the dissertation literature review. *Practical assessment, research and evaluation,* 14(1), p. 13.

Ranta Images, 2017a. *Mature beautiful woman relaxing in park Bangkok Thailand*. [Online]

Available at: https://www.istockphoto.com/es/foto/madura-hermosa-mujer-relajante-en-el-parque-de-bangkok-tailandia-gm866875720-144416237

[Accessed 24 November 2024].

Ranta Images, 2017b. *Face of handsome asian - Stock Photo*. [Online] Available at: https://www.istockphoto.com/es/foto/cara-de-asiatico-guapo-gm668887338-124816319

[Accessed 24 November 2024].

Ranta Images, 2017c. *Asian young man wearing blue shirt against city nature view - Stock Photo*. [Online]

Available at: https://www.istockphoto.com/es/foto/hombre-joven-asi%C3%A1tico-vistiendo-camisa-azul-contra-vista-de-la-naturaleza-de-la-ciudad-gm863309992-143559813

[Accessed 24 November 2024].

Rathje, S., Roozenbeek, J., Van Bavel, J. J. & van de Linden, S., 2023. Accuracy and social motivations shape judgements of (mis) information. *Nature Human Behaviour,* 7(6), pp. 892-903.

Raziel, Y., 2021. *Did ELVIS learn how to sing a JEWISH song?!*. [Online] Available at: https://www.youtube.com/watch?v=_kD7eTmj1us

[Accessed 9 March 2025].

Reality Defender, 2025. *Reality Defender Technology*. [Online] Available at: https://www.realitydefender.com/technology [Accessed 7 March 2025].

Regan, G., 2024. *A Brief History of Deepfakes*. [Online] Available at: https://www.realitydefender.com/blog/history-of-deepfakes [Accessed 24 February 2025].

Regulations on the Management of Deep Integration of Internet Information Services, 2023, Article 6. *Prohibits deepfake content that violates laws, including fake news.*. China: s.n.

Repo, S., Lehtinen, T., Rusanen, E. & Hyytinen, H., 2017. Prior education of Open University students contributes to their capability in critical thinking. *Journal of Adult and Continuing Education,* 23(1), pp. 61-77.

Resemble.AI, 2025. *Deepfake Incident Database*. [Online] Available at: https://www.resemble.ai/deepfake-database/ [Accessed 19 March 2025].

Robertson, D. J. et al., 2018. Detecting morphed massport photos: A training and individual differences approach. *Cognitive research: principles and implications,* Volume 3, pp. 1-11.

Rombach, R. et al., 2022. High-resolution image synthesis with latent diffusion models. *In Porceedings of the IEEE/CVF conference on computer vision and pattern recognition,* pp. 10684-10695.

Ryan, G., 2018. Introduction to positivism, interpretivism and critical theory. *Nurse researcher,* 25(4), pp. 41-49.

Sabir, E. et al., 2019. Recurrent convolutional strategies for face manipulation detection in videos. *Interfaces (GUI),* 3(1), pp. 80-87.

Salahdine, F. & Kaabouch, N., 2019. Social Engineering Attacks: A Survey. *Future Internet,* 11(4), p. 89.

Salahdine, F. & Kaabouch, N., 2019. Social Engineering Attacks: A Survey. *Future Internet,* 11(4), p. 89.

Samson, C., 2023. *Scammer dupes man out of $600,000 in 10 minutes by posing as friend using AI*. [Online] Available at: https://www.yahoo.com/news/scammer-dupes-man-600-000-195329300.html?ref=upstract.com&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAM_mxEnrjj8X_9w4FclqRQTE7qGLnd1MgWm0WW2PjH0Cq_bokT9_Eb8vG5Xni1DBWIbNBt2bQc0bJlJvIuNxHCzS5jmCcMdKShd_nu [Accessed 28 February 2025].

Sanders, J. G., Ueda, Y., Yoshikawa, S. & Jenkins, R., 2019. More human than human: A turning test for photographed faces. *Cognitive Research,* 4(1), p. 23.

Saxena, V., 2023. *$600,000 'Deepfake' Fraud Heats Up AI Debate in China*. [Online] Available at: https://www.asiafinancial.com/600000-deepfake-fraud-heats-up-ai-debate-in-china [Accessed 28 February 2025].

Say, G. & Vasudeva, G., 2020. Learning from digital failures? The effectiveness of firms' divestiture and management turnover responses to data breaches. *Strategy Science,* 5(2), pp. 71-145.

Schreiber, A. & Schreiber, I., 2024. Bridging knowledge gap: the contribution of employees' awareness of AI cyber risks comprehensive program to reducing emerging AI digital threats. *Information & Computer Security,* 32(5), pp. 613-635.

Schultz, 2005. The human factor in security. *Computers & Security,* Volume 24, pp. 425-426.

Sensity, 2025. *Deepfake Detection*. [Online] Available at: https://sensity.ai/deepfake-detection/ [Accessed 7 March 2025].

Seow, J. W., Lim, M. K., Phan, R. C. & Liu, J. K., 2022. A comprehensive overview of Deepfake: Generation, detection, datasets, and opportunities. *Neurocomputing,* Volume 513, pp. 351-371.

Sharma, M. & Kaur, M., 2022. A review of Deepfake technology: an emerging AI threat. *Soft Computing for Security Applications: Proceedings of ICSCS 2021,* pp. 605-619.

Shen, B. et al., 2021. A study of the Human Perception of Synthetic Faces. *In 2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021),* pp. 1-8.

Sherif, V., 2018. Evaluating preexisting qualitative research data for secondary analysis. *In Forum qualitative sozialforschung/forum: Qualitative social research,* 19(2).

Shim, K. & Yang, S. U., 2016. The effect of bad reputation: The occurrence of crisis, corporate social responsibility, and perceptions of hypocrisy and attitudes towards a company. *Public relations review,* 42(1), pp. 68-78.

Shreem Growth Partners, 2025. *Shreem Growth Partners - Business Consulting and Services*. [Online]

Available at: https://www.linkedin.com/company/shreempartners/ [Accessed 13 March 2025].

Sikra, J., Renaud, K. V. & Thomas, D. R., 2023. UK cybercrime, victims and reporting: a systematic review. *Commonwealth Cybercrime Journal,* 1(1), pp. 28-59.

Simonchik, K., 2025. *Deepfakes: What Businesses Need to Know to prevent deepfake fraud*. [Online] Available at: https://www.publicnow.com/view/EFDBC3AACA2976CF9B5D181BF9ADCE8160386960 [Accessed 28 February 2025].

Simon, F. M., Altay, S. & Mercier, H., 2023. Misinformation reloaded? Fears about the impact of generative AI on misinformation are overblown. *Harvard Kennedy School Misinformation Review,* 4(5).

Singh , P. & Dhiman, D. B., 2023. Exploding AI-Generated Deepfakes and misinformation: A threat to global concern in the 21st century. *Available at SSRN 4651093*.

Sjovall, A. M. & Talk, A. C., 2004. From actions to impressions: Cognitive attribution theory and the formation of corporate reputation. *COrporate Reputation Review,* Volume 7, pp. 269-281.

Smith, G., 2024. *Senior exec at Arup duped by deepfake vishing scam; lessons and fixes*. [Online] Available at: https://www.oryxalign.com/blog/arup-deepfake-vishing-scam [Accessed 25 February 2025].

Solo, A. M., 2025. Preventing Deepfakes From Being Used for Impersonation and Defamation. *In Deepfakes adn Their Impact on Business,* pp. 267-284.

South China Morning Post, 2025. *Hong Kongers lose B870m to scams in a week, AI voice-cloning used Please credit and share this article with others using this link: https://www.bangkokpost.com/world/2957936/hong-kongers-lose-b870m-to-scams-in-a-week-ai-voice-cloning-used. View our polic*. [Online] Available at: https://www.bangkokpost.com/world/2957936/hong-kongers-lose-b870m-to-scams-in-a-week-ai-voice-cloning-used#google_vignette

[Accessed 28 February 2025].

Southerlycourse, 2018. *Portrait of Mature Man Wearing White T-shirt stock photo*. [Online]

Available at: https://www.istockphoto.com/photo/portrait-of-mature-man-wearing-white-t-shirt-gm907524480-250009916?searchscope=image%2Cfilm

[Accessed 18 November 2024].

Spears, J. L. & Barki, H., 2010. User participation in information systems security risk management. *MIS Quaterly,* September, 34(3), pp. 503-522.

Stratton, S. J., 2024. Purposeful Sampling: Advantages and Pitfalls. *Prehospital and Disaster Medicine,* 39(2), pp. 121-122.

Stupp, C., 2019. *Voice in Unusual Cybercrime Case*. [Online] Available at: https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402

[Accessed 27 February 2025].

SumSub, 2024. *Identity Fraud Report 2024,* s.l.: SumSub.

Sun, N. et al., 2023. Cyber threat intelligence mining for proactive cybersecurity defence: A survey and new perspectives. *IEEE Communications Surveys and Tutorials,* 25(3), pp. 1748-1774.

Syaritri, W. et al., 2022. Social Engineering attacks prevention: A systematic literature review. *IEEE access,* Volume 10, pp. 39325-29343.

Sze, V., Yu-Hsin, C., Yang, T. J. & Emer, J. S., 2017. Efficient processing of deep neural networks: A tutorial and survey. *Proceedings of the IEEE ,* 105(12), pp. 2295-2329.

Taherdoost, H., 2022. What are different research approaches? Comprehensive Review of Qualitative, quantitative, and mixed method research, their applications, types, and limitations. *Journal of Management Science & Engineering Research 5,* 5(1), pp. 53-63.

Tahir, R. et al., 2021. Seeing is believing: Exploring perceptual differences in deepfake videos. *In Proceedings of the 2021 CHI conference on human factors in computing systems,* pp. 1-16.

Temir, E., 2020. Deepfake: New Era in The Age of Disinformation & End of Reliable Journalism. *Journal of Selcuk Communication,* 13(2), pp. 1009-1024.

Texas Senate Bill, 2019, sb. 751. *Relating to the creation of a criminal offense for fabricating a deceptive video with intent to influence the outcome of an election.* s.l.:s.n.

Thapa, B. & Rai, N., 2015. A study on purposive sampling method in research. *Kathmandu: Kathmandu School of Law,* 5(1), pp. 8-15.

The Indian Penal Code, 1860, s. 124. *Sedition or inciitng hatred via deepfakes.* India: s.n.

The Indian Penal Code, 1860, s. 468. *Forgery using deepfakes.* India: s.n.

The Indian Penal Code, 1860, s. 500. *Punishment for Defamation.* India: s.n.

The Indian Penal Code, 1860, s. 506. *Threats or intimidation using fabricated content.* India: s.n.

Thomas, D. R., 2003. A general inductive approach for qualitative data analysis.

Thomas, D. R., 2006. A general inductive approach for analyzing qualitative evaluation data. *American journal of evaluation,* 27(2), pp. 237-246.

Times of India, 2025. *SBI issues Public Caution Notice: Beware of.....* [Online] Available at: https://timesofindia.indiatimes.com/technology/tech-news/sbi-issues-public-caution-notice-beware-of-/articleshow/118722934.cms [Accessed 19 March 2025].

Tinwell, A., 2014. *The uncanny valley in games and animation*. s.l.:CRC press.

Tolosana, R. et al., 2020. Deepfakes and beyond: A survey of face manipulation and fake detection. *Information Fusion,* Volume 64, pp. 131-148.

Trend Micro, 2019. *Unusual CEO Fraud via Deepfake Audio Steals US$243,000 From UK Company*. [Online] Available at: https://www.trendmicro.com/vinfo/mx/security/news/cyber-attacks/unusual-ceo-fraud-via-deepfake-audio-steals-us-243-000-from-u-k-company [Accessed 27 February 2025].

TrueBees, 2025. *Can you trust what your eyes see?*. [Online] Available at: https://truebees.eu/site/ [Accessed 13 March 2025].

Truepic, 2025. *The trustworthy virtual inspection*. [Online] Available at: https://www.truepic.com/ [Accessed 7 March 2025].

Trusona, 2025. *Stop GenAI Deepfakes Attacking Your IT Helpdesk*. [Online] Available at: https://www.trusona.com/ [Accessed 7 March 2025].

Trzeciak, J. & MacKay, S. E., 1994. *Study skills for academic writing.* s.l.:Prentice Hall.

Tucciarelli, R., Vehar, N., Chandaria, S. & Tsakiris, M., 2022. On the realness of people who do not exist: The social processing of artificial faces. *Iscience,* 25(12).

Usukhbayar, B. & Homer, S., 2020. *Deepfake videos: The future of entertainment,* Berlin: Research Gate.

Vâlsan, C., Druicâ, E. & Eisenstat, E., 2022. On Deep-Fake Stock Prices and Why Investor Behaviour Might Not Matter. *Algorithms,* 15(12).

Vahl, S., 2024. *Cloned customer voice beats bank security checks*. [Online] Available at: https://www.bbc.co.uk/news/articles/c1lg3ded6j9o [Accessed 19 March 2025].

Vakulov, A., 2025. *Can Metadata Standards Like C2PA Fight Deepfakes?*. [Online] Available at: https://builtin.com/artificial-intelligence/fighting-deepfakes [Accessed 19 March 2025].

Van Den Oord, A., Kalchbrenner, N. & Koray, K., 2016. Pixel recurrent neural networks. *In International conference on machine learning,* pp. 1747-1756.

van der Nagel, E., 2020. Verifying images: deepfakes, control, and consent. *Porn Studies,* 7(4), pp. 424-429.

Van Teijingen, E. & Hundley, V., 2001. The imporance of pilot studies. *Social research update,* Volume 35, pp. 1-4.

Vecchietti, G., Liyanaarachchi, G. & Viglia, G., 2025. managing deepfakes with artificial intelligence: Introducing the business privacy calculus. *Journal of Business Research,* Volume 186, p. 115010.

Vorona, L., 2023. *Close up portrait of serious businessman, afro american man in business suit and glasses looking at camera thinking, boss outside office building stock photo*. [Online] Available at: https://www.istockphoto.com/photo/close-up-portrait-of-serious-businessman-afro-american-man-in-business-suit-and-gm1456419489-

491457822?searchscope=image%2Cfilm

[Accessed 19 November 2024].

Vosoughi, S., Roy, D. & Aral, S., 2018. The spread of true and false news online. *Science,* 359(6380), pp. 1146-1151.

Wahyu, E. R., 2023. Risk Analysis and Crisis Management: A Proactive Approach to Overcoming Business Uncertainty. *Return Study of Management Economic and Bussines,* 2(9), pp. 953-961.

Walker, D. M., 2005. Reclaiming public trust in the wake of recent corporate accountability failures. *International Journal of Disclosure and Governance,* Volume 2, pp. 264-271.

Walliman, N., 2021. *Research Methods*. 3rd Edition ed. London: Routledge.

Wang, C. et al., 2023. Neural Codec Language Models are Zero-Shot Text to Speech Synthesizers. *arXiv preprint arXiv:2301.02111*.

Wang, J., Gupta, M. & Rao, H. R., 2015. Insider threats in a financial institution. *MIS quaterly,* 39(1), pp. 91-112.

Wang, R. et al., 2020. Deepsonar: Towards effective and robust detection or ai-synthesised fake voices. *In Proceedings of the 28th ACM international conference on multimedia,* pp. 1207-1216.

Wang, T. et al., 2024, p. 25. Deepfake Detection: A Comprehensive Survey from the Reliability Perspective. *ACM Computing Surveys,* 57(3), pp. 1-35.

Waters, V. H. &. J., 2019. *Mixed Methods in Criminology*. 1st Edition ed. London: Routledge.

Webb, E., Campbell, D., Schwatrz, R. & Sechrest, I., 1966. *Unobtrusive measures*. Chicago, IL: s.n.

Weiner, B., 1974. *Achievement motivation and attribution theory*. Morristown, NJ: General Learning Press.

Weiner, B., 1986. *An attributional theory of motivation and emotion.* New York: Springer-Verlag.

Westerlund, M., 2019. The emergence of deepfake technology: A review. *Technology innovation management review,* 9(11).

Westerlund, M., 2019. The Emergence of Deepfake Technology: A Review. *Technology Innovation Management Review,* 9(11), pp. 39-52.

Whittaker, L., Letheren, K. & Mulcahy, R., 2021. The Rise of Deepfakes: A Conceptual Framework and Research Agenda for Marketing. *Australasian Marketing Journal (AMJ),* 29(3), pp. 204-214.

Whittaker, L. et al., 2023. Mapping the deepfake landscape for innovation: A multidisciplinary systematic review and future research agenda. *Technovation,* Volume 125, p. 102784.

Whittaker, L. et al., 2023. Mapping the deepfake landscape for innovation: A multidisciplinary systematic review and future research agenda. *Technovation,* Volume 125, p. 102784.

Widder, D. G., Nafus, D., Dabbish, L. & Herbsleb, J., 2022. Limits and possibilities for "Ethical AI" in open source: A study of deepfakes. *In Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency,* pp. 2035-2046.

Wilshusen, G. C. & Powner, D. A., 2009. *Cybersecurity: Continued Efforts Are Needed to Protect Information Systems From Evolving Threats,* Washington DC: https://apps.dtic.mil/sti/pdfs/ADA516401.pdf.

World Economic Forum, 2024. *This engineering firm was hit by a deepfake fraud. Here's what it learned.* [Online] Available at: https://www.weforum.org/videos/arup-deepfake-fraud/ [Accessed 25 February 2025].

Wright, K. B., 2005. Researching internet-based populations: advantages and disadvantages of online survey research, online questionnaire authorising software packages, and web survey services. *Journal of computer-mediated communication,* 10(3), p. JCMC1034.

Young, S., 2022. *How to Write Your Undergraduate Dissertation in Criminology*. Oxon: Routledge.

Yuan, C. & Cui, B., 2022. Adversarial Attack with Adaptive Gradient Variance for Deep Fake FIngerprint Detection. *In 2022 IEEE 24th International Workshop on Multimedia Signal Processing (MMSP),* pp. 1-6.

Yu, N., Skripniuk, V., Abdelnabi, S. & Fritz, M., 2021. Artificial fingerprinting for generative models: Rooting deepfake attribution in training data. *In Proceedings of the IEEE/CVF International conference on computer vision,* pp. 14448-14457.

Zekun, Y., 2023. *Police Warn of AI fraud after man duped by 'friend'*. [Online] Available at: https://www.chinadaily.com.cn/a/202305/22/WS646b4fd3a310b6054fad4731.html [Accessed 28 February 2025].

Zhang, T., 2022. Deepfake Generation and detection, a survey. *Multimedia Tools and Applications,* 81(5), pp. 6259-6276.

Zheng, Y. et al., 2021. Exploring temporal coherence for more general video face forgery detection. *In Proceedings of the IEEE/CVF international conference on computer vision,* pp. 15044-15054.

Appendix A: Full Survey Layout and Questions

The Risk of Deepfakes in Corporate Environments

Tallulah O'Hanlon- Nottingham Trent University

This survey is conducted by Tallulah O'Hanlon at Nottingham Trent University as part of the Summative assessment for Research Projects Module. The aim of this Survey is to understand how well employees in companies can detect Deepfakes.

Deepfakes are images, videos, or audio which are edited or generated using artificial intelligence tools, and which may depict real or non-existent people.

You will be presented with 36 images, where you have to decide whether you think the image is of a real person, or whether the image is a deepfake. The survey should not take more than 20 minutes to complete.

The information gathered will only be used by Tallulah O'Hanlon and may be requested by the module team. Data will be stored anonymously and securely as electric files on the NTU system and destroyed after the successful completion of the module. Results will be summarised in a dissertation and no individual involved in the survey will be identified. This survey is for individuals aged 18 and over.

Participation in this survey is voluntary. By completing and submitting this survey, you confirm that you have read and understood the above information and agree to participate in this research. If you wish to proceed,

* Required please tick the box below.

If you have any questions or concerns, please contact:

Tallulah O'Hanlon - n1021149@my.ntu.ac.uk

Tine Munk - tine.munk@ntu.ac.uk (Personal Tutor)

1

I agree that I have read and understood the information and am happy to proceed *

- ◯ I agree

2

I am over 18. *

- ◯ Yes

Characteristics

3

Unique Code Identifier *

To maintain anonymity throughout this study, please generate a unique code identifier. This code will allow you to withdraw your response at any point, if needed. For example: TLO2093

4

How old are you?

5

What would you consider your hierarchy level within your job?

- Entry Level Employee
- General Employee
- Manager
- Director
- Executive
- Board of Directors
- Other

6

What Industry do you work in?

7

What is your Gender

- ◯ Woman
- ◯ Man
- ◯ Non-binary
- ◯ Prefer not to say
- ◯ Other
- ◯ 8

What is your Ethnicity?

◯

9

What is your Highest Level of Education?

Survey

Thank you for agreeing to take part in my study of the Risk of Deepfakes in Corporate Environments. As you know this questionnaire will present various audios, videos and pictures – some real and some deepfakes- and you will be asked to identify which are genuine and which are fake. This will help explore key themes such as employee perception of deepfakes, the effectiveness of identifying them, potential risks for businesses and how organisations are preparing to counter these threats.

You do not have to answer all of the questions.

10

Is this image real or is it a deepfake?

- Real
- Deepfake
- Not sure
- Don't want to answer

11

How confident are you with your answer?

0	1	2	3	4	5	6	7	8	9	10

Not at all Confident Very Confident

12

Is this image real or is it a deepfake?

- Real
- Deepfake
- Not sure
- Don't want to answer

13

How confident are you with your answer?

0	1	2	3	4	5	6	7	8	9	10

Not at all Confident Very Confident

14

Is this image real or is it a deepfake?

- Real
- Deepfake
- Not sure
- Don't want to answer

15

How confident are you with your answer?

0	1	2	3	4	5	6	7	8	9	10

Not at all Confident Very Confident

16

Is this image real or is it a deepfake?

- Real
- Deepfake
- Not sure
- Don't want to answer

17

How confident are you with your answer?

0	1	2	3	4	5	6	7	8	9	10

Not at all Confident Very Confident

18

Is this image real or is it a deepfake?

- Real
- Deepfake
- Not sure
- Don't want to answer

19

How confident are you with your answer?

0	1	2	3	4	5	6	7	8	9	10

Not at all Confident Very Confident

20

Is this image real or is it a deepfake?

- Real
- Deepfake
- Not sure
- Don't want to answer

21

How confident are you with your answer?

0	1	2	3	4	5	6	7	8	9	10

Not at all Confident Very Confident

22

Is this image real or is it a deepfake?

- Real
- Deepfake
- Not sure
- Don't want to answer

23

How confident are you with your answer?

0	1	2	3	4	5	6	7	8	9	10

Not at all Confident Very Confident

24

Is this image real or is it a deepfake?

- Real
- Deepfake
- Not sure
- Don't want to answer

25

How confident are you with your answer?

0	1	2	3	4	5	6	7	8	9	10

Not at all Confident Very Confident

26

Is this image real or is it a deepfake?

- Real
- Deepfake
- Not sure
- Don't want to answer

27

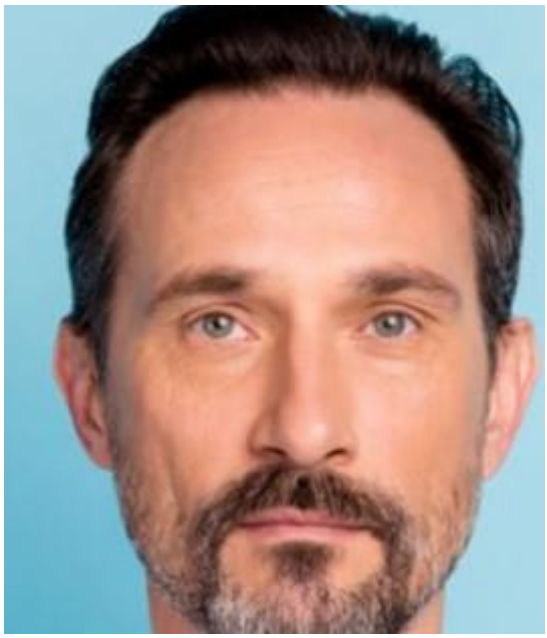
How confident are you with your answer?

0	1	2	3	4	5	6	7	8	9	10

Not at all Confident Very Confident

28

Is this image real or is it a deepfake?



- Real
- Deepfake
- Not sure
- Don't want to answer

29

How confident are you with your answer?

0	1	2	3	4	5	6	7	8	9	10

Not at all Confident Very Confident

30

Is this image real or is it a deepfake?

- Real
- Deepfake
- Not sure
- Don't want to answer

31

How confident are you with your answer?

0	1	2	3	4	5	6	7	8	9	10

Not at all Confident Very Confident

32

Is this image real or is it a deepfake?

- Real
- Deepfake
- Not sure
- Don't want to answer

33

How confident are you with your answer?

0	1	2	3	4	5	6	7	8	9	10

Not at all Confident Very Confident

34

Is this image real or is it a deepfake?

- Real
- Deepfake
- Not sure
- Don't want to answer

35

How confident are you with your answer?

0	1	2	3	4	5	6	7	8	9	10

Not at all Confident Very Confident

36

Is this image real or is it a deepfake?

- Real
- Deepfake
- Not sure
- Don't want to answer

37

How confident are you with your answer?

0	1	2	3	4	5	6	7	8	9	10

Not at all Confident Very Confident

38

Is this image real or is it a deepfake?

- Real
- Deepfake
- Not sure
- Don't want to answer

39

How confident are you with your answer?

0	1	2	3	4	5	6	7	8	9	10

Not at all Confident Very Confident

40

Is this image real or is it a deepfake?

- Real
- Deepfake
- Not sure
- Don't want to answer

41

How confident are you with your answer?

0	1	2	3	4	5	6	7	8	9	10

Not at all Confident Very Confident

42

Is this image real or is it a deepfake?

- Real
- Deepfake
- Not sure
- Don't want to answer

43

How confident are you with your answer?

0	1	2	3	4	5	6	7	8	9	10

Not at all Confident Very Confident

44

Is this image real or is it a deepfake?

- Real
- Deepfake
- Not sure
- Don't want to answer

45

How confident are you with your answer?

0	1	2	3	4	5	6	7	8	9	10

Not at all Confident Very Confident

46

Is this image real or is it a deepfake?

- Real
- Deepfake
- Not sure
- Don't want to answer

47

How confident are you with your answer?

0	1	2	3	4	5	6	7	8	9	10

Not at all Confident Very Confident

48

Is this image real or is it a deepfake?

- Real
- Deepfake
- Not sure
- Don't want to answer

49

How confident are you with your answer?

0	1	2	3	4	5	6	7	8	9	10

Not at all Confident Very Confident

50

Is this image real or is it a deepfake?

- Real
- Deepfake
- Not sure
- Don't want to answer

51

How confident are you with your answer?

0	1	2	3	4	5	6	7	8	9	10

Not at all Confident Very Confident

52

Is this image real or is it a deepfake?

- Real
- Deepfake
- Not sure
- Don't want to answer

53

How confident are you with your answer?

0	1	2	3	4	5	6	7	8	9	10

Not at all Confident Very Confident

54

Is this image real or is it a deepfake?

- Real
- Deepfake
- Not sure
- Don't want to answer

55

How confident are you with your answer?

0	1	2	3	4	5	6	7	8	9	10

Not at all Confident Very Confident

56

Is this image real or is it a deepfake?

- Real
- Deepfake
- Not sure
- Don't want to answer

57

How confident are you with your answer?

0	1	2	3	4	5	6	7	8	9	10

Not at all Confident Very Confident

58

Is this image real or is it a deepfake?

- Real
- Deepfake
- Not sure
- Don't want to answer

59

How confident are you with your answer?

0	1	2	3	4	5	6	7	8	9	10

Not at all Confident Very Confident

60

Is this
image real
or is it a
deepfake?

- Real
- Deepfake
- Not sure
- Don't want to answer

61

How confident are you with your answer?

0	1	2	3	4	5	6	7	8	9	10

Not at all Confident Very Confident

62

Is this image real or is it a deepfake?

- Real
- Deepfake
- Not sure
- Don't want to answer

63

How confident are you with your answer?

0	1	2	3	4	5	6	7	8	9	10

Not at all Confident Very Confident

64

Is this image real or is it a deepfake?

- Real
- Deepfake
- Not sure
- Don't want to answer

65

How confident are you with your answer?

0	1	2	3	4	5	6	7	8	9	10

Not at all Confident Very Confident

66

Is this image real or is it a deepfake?

- Real
- Deepfake
- Not sure
- Don't want to answer

67

How confident are you with your answer?

0	1	2	3	4	5	6	7	8	9	10

Not at all Confident Very Confident

68

Is this image real or is it a deepfake?

- Real
- Deepfake
- Not sure
- Don't want to answer

69

How confident are you with your answer?

0	1	2	3	4	5	6	7	8	9	10

Not at all Confident Very Confident

70

Is this image real or is it a deepfake?

- Real
- Deepfake
- Not sure
- Don't want to answer

71

How confident are you with your answer?

0	1	2	3	4	5	6	7	8	9	10

Not at all Confident Very Confident

72

Is this image real or is it a deepfake?

- Real
- Deepfake
- Not sure
- Don't want to answer

73

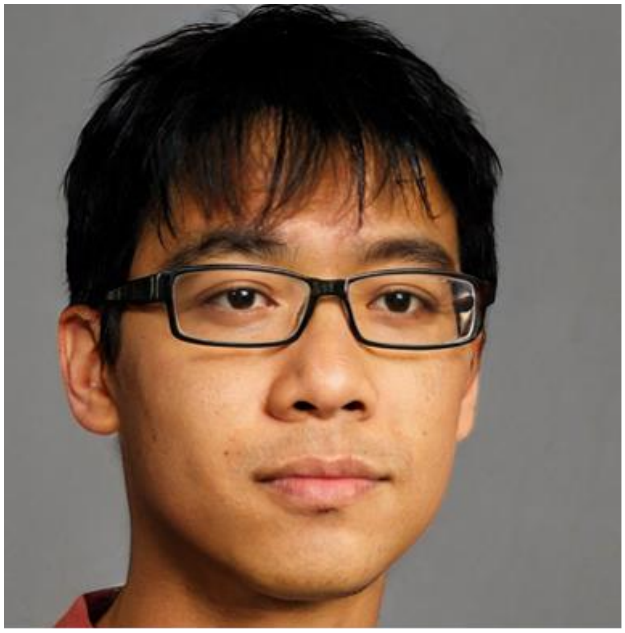
How confident are you with your answer?

0	1	2	3	4	5	6	7	8	9	10

Not at all Confident Very Confident

74

Is this image real or is it a deepfake?



- Real
- Deepfake
- Not sure
- Don't want to answer

75

How confident are you with your answer?

0	1	2	3	4	5	6	7	8	9	10

Not at all Confident Very Confident

76

Is this image real or is it a deepfake?

- Real
- Deepfake
- Not sure
- Don't want to answer

77

How confident are you with your answer?

0	1	2	3	4	5	6	7	8	9	10

Not at all Confident Very Confident

78

Is this image real or is it a deepfake?

- Real
- Deepfake
- Not sure
- Don't want to answer

79

How confident are you with your answer?

0	1	2	3	4	5	6	7	8	9	10

Not at all Confident Very Confident

80

Is this image real or is it a deepfake?

- Real
- Deepfake
- Not sure
- Don't want to answer

81

How confident are you with your answer?

0	1	2	3	4	5	6	7	8	9	10

Not at all Confident Very Confident

Debrief

Thank you for taking the time to participate in this study. This document provides background about the research and more information on why I am conducting this study.

You have just participated in a research study conducted by Tallulah O'Hanlon my contact information can be found below:
Email:
N1021149@my.ntu.ac.uk

I am an undergraduate student on the BA (Hons) Criminology course in the School of Social Sciences. The purpose of this study was to explore the risks of deepfakes in corporate environments through analysing well employees can detect deepfakes. As you are aware from the information document, your participation in this study is voluntary. If you wish you may withdraw your data at any time up to 3 weeks after taking part in this study, at which point all records of your involvement in the research will be deleted. Please make sure you have made a note of your unique identifier in case you need to contact us. If you wish to withdraw you do not need to provide reasoning. To withdraw you will need to contact either me, Tallulah (N1021149@my.ntu.ac.uk) or my project supervisor Tine (Tine.Munk@ntu.ac.uk) Please note that you may lose anonymity if you contact me via an email address that exposes your identity, or if your correspondence reveals your identity in any way. Nevertheless, any requests of withdrawal will remain private and confidential, and any data held will be destroyed securely. Please also be aware that confidentiality will be broken if you disclose anything of a criminal nature or anything that puts you or anyone else in danger. All retained data will be kept in raw form until processed for analysis before being destroyed. Only I will have access to the raw data. My supervisor and I will have access to the processed anonymised data Please keep a copy of this debrief form for your records

---

Appendix B: Consent Form

Agreement To Consent

My Unique ID is:…….
(please avoid any self-identifying information e.g., real names or birthdates)

	Tick box to confirm.
I confirm that I have read and understood the purpose of this research what the experiment entails and what my role in this research includes.	
I confirm that I am aged 18 years or older	
I confirm that I understand what will happen with all my data once they have been recorded	
I confirm that I understand how my data will be stored and handled	
I give consent for my interview to be recorded and transcribed.	
I confirm that I will provide a unique identification number in case I need to contact the researchers about my participation	
I confirm that I understand how to contact the researchers to ask questions, raise concern or withdraw my data, using my unique identification number and what will happen if I do	
I confirm that I understand that I have the right to withdraw my data at any point during or after taking part up until two weeks after today's date	
I confirm that having read and confirmed the above I freely volunteer to take part in this study	

Appendix C: Interview Schedule

Interview Schedule

Semi-structured Interview Schedule for the study of Risks of Deepfakes in Corporate Environments

Introduction

Thank you for agreeing to take part in my study of the Risk of Deepfakes in Corporate Environments. As you know we will be discussing concepts such as public perception, current and future threats of deepfakes, corporate preparedness and response measures, and measures put in place for companies. If you wish to stop the interview at any time, please do let the researcher know.

- How aware do you think employees and executives are of the potential risks deepfakes pose to the company?
- Do you think public awareness of deepfakes has any impact on corporate security? Why or why not?
- Have you encountered any misconceptions about deepfakes that could affect how seriously companies take the threat?
- How do you think public perception of deepfakes influences policymaking or regulations for corporate cybersecurity?
- Have you observed a change in how clients or customers view deepfake risks over the last few years?

- What are some current examples of deepfake use that you think companies should be aware of?
- In what ways do you anticipate deepfake technology being used to target employees or company executives?
- Are there particular departments or functions within companies that are more vulnerable to deepfake attacks?
- Are there emerging signs or trends that point to how deepfakes could become more prevalent in social engineering attacks?
- What is your opinion on the level of corporate preparedness across industries when it comes to deepfakes?
- Do you think there are any gaps in current corporate response strategies for deepfake threats?
- How well do corporate response plans for deepfakes align with those for other cybersecurity threats, like phishing or ransomware?
- What factors do you think drive the level of preparedness a company has for deepfake-related threats?
- What risks do deepfakes pose to the internal culture or employee morale within a company?
- How do you think company image can be impacted by deepfakes?
- How do you think deepfakes could affect relationships with investors or stakeholders?
- What are the potential legal implications for companies impacted by deepfake incidents?
- In what ways could deepfakes be used to manipulate or disrupt the supply chain within an industry?

- How frequently do you think companies should be updating their technology to stay on top of deepfake threats?
- Are there any innovative technologies or tools you have found particularly effective against deepfakes?
- How can companies balance the cost of implementing deepfake countermeasures with the potential risks?
- What kind of training or awareness programs do you believe are essential for employees to recognise and respond to deepfake threats?

Thank you for taking time to be a part of this study. If you need any support after this interview, please see the debrief form. The debrief document also expresses how you can request to withdraw from the research if you no longer feel comfortable taking part in this study.

If you do have any questions or queries about the research, please do not hesitate to get in touch via the details below.

Researchers Email: N1021149@my.ntu.ac.uk

Supervisor Email: tine.munk@ntu.ac.uk

School of Social Sciences

Nottingham Trent University

50 Shakespeare Street

Nottingham NG1 4FQ

Appendix D: Participant Information Sheet

Informed Consent Form to Participate in Criminological Research

The purpose of this study is to explore the risks of deepfakes in corporate environments through analysing how companies are preventing deepfakes, the current and future extents of deepfakes and public perceptions. This study is part of a third-year research project at the Nottingham Trent University, in the Criminology and Criminal Justice Department.

Procedure

You will take part in a semi-structured interview to discuss your experience of working in the Cybersecurity field and how learning about the evolving landscape has shaped the way you think. At the start, you'll be asked to share some background information, including your role and level of experience with deepfake threats. The interview will explore several key themes to guide our conversation: public perception of deepfakes, current and future deepfake threats, how companies are preparing for these risks and response measures, potential risks for businesses, and the specific measures organisations are using to counter these risks. If you need to pause or withdraw from the interview at any time, please inform the researcher.

These interviews will take place online Via Microsoft Teams and will last approximately 45 minutes. The interview will be recorded and transcribed via the sound recorder on I, the researcher's laptop.

Anonymity

The laptop used for recording the interviews is password protected therefore ensuring all raw data is secure and all data will be stored according to GDPR and UK Data Protection Act 2018 regulations in an anonymised format. Once analysis has taken place and completed the recordings will be destroyed. Only I, the researcher and my supervisor will have access to the data.

Every effort will be made to maintain your anonymity when collecting and representing the data from the interviews. The analysed and processed data will appear in any submissions or publications. However please be aware if anything of a criminal nature or anything that puts you in danger or anyone else in danger is disclosed this will have to be reported to the relevant individuals and therefore confidentiality will be broken.

Withdrawal from the research

In order to protect your identity, you will be given a unique identifier and all identifiable/revealing information regarding any incidents will be changed or retracted from the study. If you wish to withdraw your data from the study, you will have to provide your unique identifier and you have the right to do this up until 3 weeks after today's date without any reason or explanation.

If you should wish to withdraw, please contact the researcher and my supervisor and ask for your data to be withdrawn. Please be aware that you risk losing your anonymity if the method of contact or correspondence reveals your identity. However, only the researcher and supervisor will have access to this withdrawal request, and this will be deleted along with all requested data.

All participation in this research is completely voluntary, please make sure you understand the nature of this research prior to the interview process. If you wish to proceed with this research your participation is greatly appreciated. Please complete the following consent form. If you have any questions or concerns before, during or after the research has been conducted then please find my contact details and those of my supervisor at the bottom of this form.
Thank you for agreeing to consider participation in this research project.

Investigators Contact Details:

Tallulah O'Hanlon (BA (Hons) Criminology)
Email: N1021149@my.ntu.ac.uk

Supervisor: Tine Munk
Email: tine.munk@ntu.ac.uk
Tel: +44 115 84 85544
School of Social Sciences
Nottingham Trent University
50 Shakespeare Street
Nottingham NG1 4FQ

Appendix E: Debrief Form

Debrief Form for Participants

Thank you for taking the time to participate in this study. This document provides background about the research and more information on why I am conducting this study.

You have just participated in a research study conducted by **Tallulah O'Hanlon** my contact information can be found below:

Email: N1021149@my.ntu.ac.uk

I am an undergraduate student on the BA (Hons) Criminology course in the School of Social Sciences.

The purpose of this study was to explore the risks of deepfakes in corporate environments through analysing how companies are preventing deepfakes, the current and future extents of deepfakes and public perceptions.

As you are aware from the information document, your participation in this study is voluntary. If you wish you may withdraw your data at any time up to 3 weeks after taking part in this study, at which point all records of your involvement in the research will be deleted. Please make sure you have made a note of your unique identifier in case you need to contact us. If you wish to withdraw you do not need to provide reasoning.

To withdraw you will need to contact either me, Tallulah (N1021149@my.ntu.ac.uk) or my project supervisor Tine (Tine.Munk@ntu.ac.uk)

Please note that you may lose anonymity if you contact me via an email address that exposes your identity, or if your correspondence reveals your identity in any way. Nevertheless, any requests of withdrawal will remain private and confidential, and any data held will be destroyed securely. Please also be aware that confidentiality will be broken if you disclose anything of a criminal nature or anything that puts you or anyone else in danger.

All retained data will be kept in raw form until processed for analysis before being destroyed. Only I will have access to the raw data. My supervisor and I will have access to the processed anonymised data.

Please keep a copy of this debrief form for your records and in case of any further questions or concerns. If you do have any questions or queries about the research, please do not hesitate to get in touch via the details below.

Researchers Email: N1021149@my.ntu.ac.uk

Supervisor Email: tine.munk@ntu.ac.uk

School of Social Sciences
Nottingham Trent University
50 Shakespeare Street
Nottingham NG1 4FQ

Should you require any additional support before, during or after this study please contact:

- Nottingham Women's Centre - Nottingham Women's Centre - Run by women for women (nottinghamwomenscentre.com)

- The Samaritans - Samaritans | Every life lost to suicide is a tragedy | Here to listen

  Call: 116 123 for free

  Email: jo@samaritans.org

- Mind – Blue Light Infoline

  Call: 0300 303 5999

  Email: Bluelightinfo@mind.org.uk

  Text: 84999

  Website: Home - Mind

- Nottingham Mental Health Helpline

  Call: 0808 196 3779

  Website: Nottingham Mental Health Helpline | Turning Point (turning-point.co.uk)

Appendix F: Survey Images

MS Forms Questionnaire for the study of Risks of Deepfakes in Corporate Environments

Introduction

Thank you for agreeing to take part in my study of the Risk of Deepfakes in Corporate Environments. As you know this questionnaire will present various audios, videos and pictures – some real and some deepfakes- and you will be asked to identify which are genuine and which are fake. This will help explore key themes such as employee perception of deepfakes, the effectiveness of identifying them, potential risks for businesses and how organisations are preparing to counter these threats.

Link to form Below

https://forms.office.com/e/eZ4N02P7y1

Pictures

White Male - Deepfakes

1. (Karras, et al., 2024)
2. (Karras, et al., 2024)

3. (Karras, et al., 2024)

White Male - Real

a. (Izusek, 2023)

b. (Deagreez, 2020)

c. (Southerlycourse, 2018)

White Female – Deepfakes

1. (Karras, et al., 2024)

2. (Karras, et al., 2024)

3. (Karras, et al., 2024)

White Female – Real

a. (Alvarez, 2024)

b. (Photique, 2014)

c. (Magann, 2019)

Black Male – Deepfakes

1, (Karras, et al., 2024)

2. (Karras, et al., 2024)

3. (Karras, et al., 2024)

Black Male - Real

1. (Lemon Photo, 2024)
2. (moneybusinessimages, 2024)
3. (Vorona, 2023)

Black Female – Deepfakes

1. _(Generated Photos, 2024)
2. (Generated Photos, 2024)
3. (Generated Photos, 2024)

Black Female - Real

1. (leezsnow, 2005)
2. (Cecillie_Arcurs, 2017)
3. (moneybusinessimages, 2015)

Asian Male – Deepfakes

1. (Karras, et al., 2024)

2. (Karras, et al., 2024)

3. (Generated Photos, 2024)

Asian Male - Real

1. (Ranta Images, 2017c)
2. (Ranta Images, 2017b)
3. (leungchopan, 2016)

Asian Female – Deepfakes

1. (Karras, et al., 2024)

2. (Karras, et al., 2024)

3. (Karras, et al., 2024)

Asian Female – Real

1. (feellife, 2013)
2. (Ranta Images, 2017a)
3. (PacoRomero, 2012)

Appendix G: Ethics Approval

Dear student,

The Social Sciences Research Ethics Committee have provided the following opinion in respect of your research ethics application:

Decision:

Favourable Ethics Opinion – You have received a favourable ethics opinion. When you commence your research activity, it should be in-line with the application made and in the knowledge that it is your responsibility to ensure that your research does not breach the integrity of the application.

We recommend that students should carefully check all participant facing documentation for typographical, grammatical and/or punctuation errors to ensure clarity of understanding.

If at any stage of the application process it has been decided that your project requires a **Disclosure and Barring Service Check** (DBS Check) or an Overseas Police Check you may **not** commence research until this check has been completed and considered as satisfactory. Please note a DBS check might not be listed as an additional condition/recommendation identified by SREC as we might be satisfied that your Project Supervisor has already identified this as a requirement on your application form.

Further information and guidance can be found on the ethics module (XXSOC10002) on NOW.

If you have any queries, please do not hesitate to contact your project supervisor or alternatively e-mail SOC.ethics@ntu.ac.uk.

Kind Regards,

Kate Bradley
(on behalf of SOC Ethics)
School Administrator School of Social Sciences
Nottingham Trent University, 50 Shakespeare Street, Nottingham, NG1 4FQ

Appendix H: User Comments on TikTok

Username	Comment	Date
Carl Christian	😂😂😂😂😂	06-01-2025
CA CA	Right down to the diapers! Priceless!!	10-02-2025
joevera120	😂😂😂😂😂 this is too much for me, hahaha..., can you make a video of me slapping Trump on the face	10-02-2025
CinePark-AI	😂😂😂😂😂	24-01-2025
Annette Engström	😂😂😂😂😂	29-12-2024
Noah Ingham	This is funny as hell 😂😂😂	23-01-2025
Nezig101	Hahahah nice	26-12-2024
Lucyp537	This is equal parts terrifying and hilarious 😂😂😂	30-01-2025
clownplonk	please I need AI movies like this 😂😂😂	30-01-2025
Fun Pop	very funny edition 😂😂😂	25-02-2025
Milton Freire Sport	😂😂😂😂😂	19-01-2025
OliviaC99	😂😂😂😂😂	31-01-2025
franka_28_5	😂😂😂😂😂	02-01-2025
Ibrahim	😂😂😂😂😂	21-01-2025
x5WOLK	This is the best 😂😂😂	21-01-2025
Doug the hand puppet	lol 😂	03-01-2025
zoppetti_cristian	😂😂😂😂😂	17-12-2024
BellaAzuqueltaplace	😂😂😂😂😂	19-12-2024
Bettina Branch. moki	😂😂😂😂😂	15-11-2024
RebelG143	😂😂😂😂😂	17-11-2024

Funny Videos World	😂😂😂😂😂	30-12-2024
Mr T	😂😂😂	30-01-2025
Raven Claw	I'm dying 😂😂😂	30-01-2025
stuffsoreon	WTF???? 😂😂😂🔥🔥	30-01-2025
Ri	I am cackling 😂😂😂😂😂	30-01-2025
DarthVader'sDaughter	best thing I've ever seen 😂😂😂	18-11-2024
Edna Thompson	😂😂😂😂😂	18-11-2024

Appendix I: Reported Company Deepfake Incidents from 2024-2025, Researchers own adapted from Resemble.AI (Resemble.AI, 2025)

Reported Company Deepfake incidents				
Date	**Incident**	**Target**	**Type**	**Source**
Sep 18, 2024	**Hong Kong Police Fraud:** Fraudsters used a deepfake video call to impersonate company officers, tricking an employee into transferring HK$200 million (£16M). AI-generated voices were likely used.	Hong Kong company employee	Audio Video	(Leng & Ho-him. 2024)
Nov 29, 2024	**BBC Voice ID Bypass:** A BBC journalist used AI voice cloning to access Santander and Halifax bank accounts, exposing weaknesses in biometric security.	Santander and Halifax banks	Audio	(CX Today, 2024)
Nov 27, 2024	**BBC Deepfake Bank Access:** A BBC reporter accessed their bank accounts using an AI-generated voice clone, bypassing security with basic audio tools.	Santander and Halifax bank accounts	Audio	(Vahl, 2024)
Mar 13, 2025	Deepfake videos of experts from The Baker Heart and Diabetes Institute promoting a diabetes supplement; deepfake images of Dr. Karl Kruszelnicki used on Facebook to sell pills; TikTok Shop sellers manipulated doctors' videos to falsely endorse products.	The Baker Heart and Diabetes Institute experts, Dr. Karl Kruszelnicki, and other doctors	Image Video Audio	(Given, 2025)
Mar 10, 2025	**Sony Music Deepfakes:** Over 75,000 AI-generated deepfakes of Sony artists were removed. Sony opposes the UK's proposed 'opt-out' AI training system.	Sony Music artists	Audio Video	(Dredge, 2025)
Mar 05, 2025	**SBI Investment Scam:** Deepfake videos falsely showed SBI executives endorsing investment schemes. The bank denied any involvement.	State Bank of India (SBI) and its customers	Video Audio	(The Times of India, 2025)
Jan 16, 2025	**£27M Deepfake Fraud:** Fraudsters stole £27M using deepfake audio and forged emails.	a UAE company	Audio	(Vakulov, 2025)

Appendix J: Countries where Deepfakes were reported March 2024- March 2025 (Resemble.AI, 2025)

Country	Number of Deepfakes
Unknown	131
UK	25
USA	17
India	7
China	4
France	3
Russia	3
Australia	2
Hong Kong	2
Canada	1
Spain	1
Abu Dhabi	1
Indonesia	1

Appendix K: One-sided T-Test

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Deepfake Detection Accuracy	229	32.3629	11.33434	.74899

One-Sample Test

Test Value = 53.16

	t	df	Significance One-Sided p	Significance Two-Sided p	Mean Difference	95% Confidence Interval of the Difference Lower	95% Confidence Interval of the Difference Upper
Deepfake Detection Accuracy	-27.767	228	<.001	<.001	-20.79707	-22.2729	-19.3212

One-Sample Effect Sizes

		Standardizer[a]	Point Estimate	95% Confidence Interval Lower	95% Confidence Interval Upper
Deepfake Detection Accuracy	Cohen's d	11.33434	-1.835	-2.046	-1.622
	Hedges' correction	11.37179	-1.829	-2.040	-1.616

a. The denominator used in estimating the effect sizes.
Cohen's d uses the sample standard deviation.
Hedges' correction uses the sample standard deviation, plus a correction factor.

Appendix L: Paired Samples T-Test

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Real_Accuracy	34.570596797	229	17.318483423	1.1444375977
	Deepfake_Accuracy	30.155264435	229	16.948802428	1.1200083900

Paired Samples Correlations

		N	Correlation	Significance One-Sided p	Significance Two-Sided p
Pair 1	Real_Accuracy & Deepfake_Accuracy	229	-.125	.030	.059

Paired Samples Test

		Paired Differences					t	df	Significance One-Sided p	Significance Two-Sided p
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper				
Pair 1	Real_Accuracy - Deepfake_Accuracy	4.4153323626	25.700479276	1.6983354746	1.0688927511	7.7617719741	2.600	228	.005	.010

Paired Samples Effect Sizes

			Standardizer[a]	Point Estimate	95% Confidence Interval Lower	95% Confidence Interval Upper
Pair 1	Real_Accuracy - Deepfake_Accuracy	Cohen's d	25.700479276	.172	.041	.302
		Hedges' correction	25.785408356	.171	.041	.301

a. The denominator used in estimating the effect sizes.
Cohen's d uses the sample standard deviation of the mean difference.
Hedges' correction uses the sample standard deviation of the mean difference, plus a correction factor.