UNIVERSITY OF TURKU

Under pressure

The influence of psychological and cognitive factors in deepfake detection under time pressure

Information Systems Science

Master's thesis

Author(s):

Sten van Dijk

Supervisor(s):

Dr. Farhan Ahmad

Dr. Emiel Caron

31.07.2025

Rotterdam

The originality of this thesis has been checked in accordance with the University of Turku quality assurance system using the Turnitin Originality Check service.

Master's thesis

Subject: Name
Author(s): Sten van Dijk
Title: Under pressure: the influence of psychological and cognitive factors in deepfake detection under time pressure
Supervisor(s): Dr. Farhan Ahmad & Dr. Emiel Caron
Number of pages: 71 pages + appendices 11 pages
Date: 31.07.2025

Style of the abstract is **Abstract**.

Keywords: deepfake, deepfake detection, cybersecurity, human vulnerabilities, and social engineering

Abstract

Advances in AI have produced increasingly realistic deepfakes, which are exploited in social engineering attacks. Studies show human ability to detect these fake videos is generally poor and varies widely (accuracy 23–87%). Time pressure further undermines performance by depleting cognitive resources and encouraging heuristic processing. This thesis examines psychological and cognitive factors affecting deepfake detection under time constraints. Specifically, it investigates how Big Five personality traits, sense of coherence (stress resilience), cognitive workload, and prior deepfake detection experience influence accuracy in a time-pressured deepfake video evaluation task. A quantitative explanatory design with 89 participants was used. Participants completed measures of personality (BFI-10) and sense of coherence (SOC-13), reported cognitive workload (NASA-TLX), and then performed a timed deepfake detection task adapted from Köbis et al. (2021). Each participant viewed ten short videos (five real, five deepfake), only once, under time pressure, and judged each as real or fake. Multiple regression and fuzzy-set qualitative comparative analysis (fsQCA) assessed how these factors predicted detection accuracy.

Regression analysis revealed that only prior detection experience significantly predicted accuracy ($\beta \approx .04$, $p = .001$), whereas no personality trait, sense of coherence score, or workload rating had a significant effect. FsQCA identified multiple distinct configurations of conditions associated with both high and low detection accuracy, underscoring the principle of equifinality in performance outcomes. Notably, the personality traits openness to experience and neuroticism emerged as important contributors within several configurations. Notably, participants were far more accurate in identifying real videos (≈75% correct) than deepfakes (≈48% correct), reflecting a truth-bias under time pressure. These findings suggest that personality, cognitive workload, and general stress resilience play a more nuanced role in deepfake detection under pressure, whereas hands-on experience is critical. Accordingly, organizations should implement targeted training interventions tailored to employees to improve their ability to detect deepfakes under time constraints. Such focused training can complement technical defences and strengthen organizational resilience to deepfake-based social engineering.

Recommendations: Develop and deploy tailored training programs in organizational settings that build deepfake detection skills under pressure, adapting content to employee experience to enhance real-world detection accuracy.

Acknowledgements

If I have seen further, it is by standing on the shoulders of giants – Isaac Newton

First and foremost, I would like to express my deepest gratitude to my thesis supervisor, Dr. Farhan Ahmad. Without his expert guidance and insightful feedback, the work you are about to read would not have reached its current quality. Dr. Ahmad supported me with both expertise and kindness throughout every stage of this research, for which I am sincerely thankful.

I am also grateful to the members of my thesis seminar committee, Dr. Emiel Caron and Prof. Hannu Salmela, whose input helped refine both the theoretical framework and methodological design. Their thoughtful questions challenged me to think more critically and significantly strengthened the quality of this work. I hope that the multi-institutional support, from both the University of Turku and Tilburg University, fostered through the double master's program, has further enriched the final result.

This research would not have been possible without the support of my fellow students in the Information Technology for Enterprise Management program. Special thanks go to Nick Roos, Pepijn Verhoef, and Sean Bekkers for their academic support through discussions and feedback, as well as for their emotional support during our time in Turku, Finland.

Lastly, I would like to thank all the participants who generously contributed their time to complete the deepfake detection experiments and questionnaires. Their willingness to engage in this study was essential to deepening our understanding of the human factors that influence detection performance under time pressure.

Sten van Dijk

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 Introduction

The introduction of this thesis provides the foundation for the research and is divided into five sections. It begins with the background and problem statement, followed by the main research question. The relevance of the study is then discussed, highlighting its academic and practical importance. Next, the research design is outlined, including the methodology and data collection. Finally, the structure of the thesis is presented, offering a roadmap for the chapters that follow.

1.1 Background/problem statement

Since the widespread adoption of the internet, reliance on information technology (IT) has grown significantly, both in society and within organizations (Cascio & Montealegre, 2016). Within online interactions, audio-visual content plays a dominant role. Advances in deep learning and artificial intelligence have further improved the quality of synthetic media, including deepfakes, a subset of synthetic media that manipulates or generates media using DL-based approaches (Groh et al., 2021; Altuncu et al., 2022). Deepfakes are becoming increasingly realistic due to ongoing advancements in AI, making the ability to distinguish real videos from deepfakes crucial (Westerlund, 2019; Chakraborty & Naskar, 2024). Human susceptibility to deepfake deception is a growing concern. Studies show significant variation in human accuracy when detecting deepfake videos, with human detection rates ranging from 23% to 87% (Bray et al., 2023). Cybercriminals exploit these vulnerabilities with social engineering attacks, such as deepfake-based payment fraud (Bateman, 2020). These attacks impersonate high-ranking employees to deceive staff into transferring sensitive data or financial assets (Rancourt & Smaili, 2023; Dsouza, 2024). Real-world examples include a finance worker in Hong Kong being tricked into sending $25 million after interacting with a deepfake impersonating their CFO and scammers stealing $20 million by impersonating Arup's CFO using deepfake technology (CNN, 2024; Guardian, 2024). As cybercriminals refine their tactics, understanding human susceptibility to deception becomes critical for organizational security.

Academic research has primarily focused on technological tools for detecting deepfakes, which demonstrate high reliability with accuracy rates between 90% and 95% (Wang et al., 2024). However, these systems are not foolproof and may be bypassed by sophisticated attackers (Kapoor & Rahman, 2024). Moreover, off-the-shelf detection systems are not widely implemented across organizations, leaving gaps in real-world applicability. Human vigilance remains essential for detecting deepfakes in practical scenarios. Time pressure significantly influences human ability to

detect deepfakes. Studies show that performing tasks under time constraints reduces cognitive resources, increasing reliance on heuristics and peripheral processing rather than critical evaluation (Goh, 2023; Chowdhury et al., 2020). This makes individuals more susceptible to deception. In deepfake-based attacks, cybercriminals could exploit urgency and time-sensitive situations to manipulate employees into making quick decisions without thoroughly evaluating the authenticity of the media presented. Human vulnerabilities, which are rooted in psychology and behavior, are central to social engineering attacks, as they are exploited through manipulation and deception rather than technical means (Dsouza et al., 2024). Psychological triggers like cognitive overload, authority, and reciprocation (Gragg, 2003) make individuals more susceptible. Recent research highlights, for example, that cognitive workload, digital literacy, and personality traits significantly influence susceptibility to social engineering attacks, although findings about personality remain inconclusive (Montanez et al., 2020; Alsobeh et al., 2024; Wang et al., 2021).

While prior studies have explored psychological and cognitive biases exploited by social engineering attacks, such as overconfidence and reduced attention under pressure, there is limited research specifically addressing human-based differences in detecting deepfakes under time constraints (Xu et al., 2022; Chowdhury et al., 2020). Research on phishing attacks provides valuable insights, demonstrating how attackers extensively use urgent cues, like "Immediate action required," to create artificial time pressure. Studies by Butavicius et al. (2022) illustrate that such urgency cues majorly increase phishing success rates by reducing critical evaluation and increasing reliance on heuristics. This established pattern in phishing research suggests similar mechanisms may operate in deepfake detection scenarios where time pressure is present. However, the specific cognitive and psychological factors that mediate this effect in deepfake detection remain unexplored.

Despite advancements in technological detection systems and insights into social engineering mechanisms, there is insufficient understanding of how cognitive and psychological factors influence employees' ability to detect deepfakes under time pressure. Psychological traits such as personality characteristics and sense of coherence could play a role in susceptibility to deception, but remain underexplored in the context of deepfake detection. This study aims to address this gap by investigating the cognitive and psychological factors that affect employees' ability to detect deepfakes under time constraints. Understanding these factors is essential for developing effective mitigation strategies to minimize the impact of deepfake-based attacks in organizational settings, leading to the following research question: What cognitive and psychological factors impact employees' deepfake detection accuracy under time pressure?

1.2 Research question

This study seeks to answer the following central research question:

‘’What cognitive and psychological factors influence employees' deepfake detection accuracy under time pressure?’’

The aim is to explore how elements such as cognitive workload, stress, personality traits, and detection experience influence an individual’s ability to identify deepfakes when operating under limited time accurately. By focusing on human-centered detection in time-pressured contexts, the research addresses a critical gap in the intersection of cybersecurity, psychology, organizational behavior, and social engineering.

1.3 Research relevance

Understanding the cognitive and psychological factors affecting employees' ability to detect deepfakes under time pressure is important for both businesses and academia. For businesses, this research helps mitigate financial risks from deepfake-based fraud, such as Arup’s or Hong Kong's example. It enhances cybersecurity strategies by integrating human vigilance alongside technological detection systems. It also aids in protecting organizational reputation by improving employee responses to deepfake threats. The research addresses a critical gap in current security approaches, as human error accounts for 95% of cybersecurity breaches (Triplett, 2022).

For academia, the study contributes to social engineering literature by incorporating theories like the Elaboration Likelihood Model while advancing understanding of cognitive vulnerabilities in digital deception contexts. It explores human vulnerabilities in deepfake detection, providing insights into the impact of personality traits that could have broader implications for cybersecurity research. These findings can inform organizational interventions and contribute to the emerging field of cyberpsychology, which examines how psychological factors influence cybersecurity behaviors in organizational settings. The research bridges the gap between theoretical frameworks and practical applications, addressing calls from practitioners for more human-centric approaches to cybersecurity that move beyond technical solutions.

Overall, the research has the potential to improve organizational security and contribute to academic fields like cybersecurity, psychology, organizational behavior, and social engineering

1.4 Research design

This thesis employs a quantitative explanatory design to investigate the cognitive and psychological factors that influence employees' deepfake detection accuracy under time pressure. The study employs a dual analytical approach, combining multiple regression analysis with fuzzy-set Qualitative Comparative Analysis (fsQCA) to offer both symmetric and asymmetric insights into the relationships between predictor variables and deepfake detection performance.

The research methodology centers on a cross-sectional survey design incorporating standardized psychological instruments and a controlled deepfake detection task. Data collection involves three phases: pre-task psychological assessment using validated scales (BFI-10, SOC-13), a deepfake detection task based on Köbis et al. (2021), and post-task cognitive workload evaluation using the NASA-TLX. This approach allows for a comprehensive examination of how personality traits, sense of coherence, cognitive workload, and prior detection experience interact to influence detection accuracy under realistic time constraints.

1.5 Research structure

The structure of this thesis follows a logical progression that supports a thorough investigation of the research question: *"What cognitive and psychological factors influence employees' deepfake detection accuracy under time pressure?"* Chapter 1 introduces the research by outlining the problem statement, objectives, and theoretical relevance. It situates deepfake threats within organizational contexts and identifies a gap in understanding human-centered detection under time constraints. Chapter 2 reviews the literature on deepfake technology, social engineering, and psychological vulnerabilities, contrasting human and technological detection capabilities. It also introduces the Elaboration Likelihood Model as the theoretical lens and explores empirical insights into personality traits, stress, and cognitive workload. Chapter 3 presents the quantitative explanatory research design, explaining the dual analytical approach, multiple regression, and fsQCA, alongside the instruments, sampling strategy, and ethical considerations. Chapter 4 reports the results, including descriptive statistics, regression outcomes, and fsQCA configurations, revealing both linear and configurational patterns that influence detection performance. Chapter 5 interprets the findings within the conceptual framework, highlights theoretical and practical implications, reflects on study limitations, and outlines future research directions. Finally, Chapter 6 provides a concise conclusion to the thesis. This structure ensures a clear and coherent path from conceptual grounding to actionable insights.

2 Literature review

This chapter provides an overview of the current scientific literature on deepfakes, with a particular focus on their use in social engineering and deepfake-based payment fraud. It begins by defining deepfakes and outlining the main types and technological developments in the field. The review then examines the risks that deepfakes pose to organizations and individuals, followed by a discussion of the challenges in detecting deepfakes, both by humans and machines. The chapter further explores how deepfakes are integrated into social engineering attacks, highlighting key theoretical frameworks and identifying human vulnerabilities that are exploited in these scenarios. In doing so, particular attention is given to the cognitive and psychological factors that influence individuals' ability to detect deepfakes under time pressure. This comprehensive review establishes the background and context for the empirical research presented in this thesis. Throughout the literature review, important gaps in the current literature are acknowledged. Finally, the conceptual model of this research is presented.

2.1 Deepfakes

Deepfakes are digitally manipulated or synthetically generated media, encompassing images, audio, and video, produced through advanced deep learning methods. Common categories of deepfake types include fake speech forgeries, fake face images, and fake videos that combine both (Altuncu, 2022). Van der Sloot & Wagens (2022) conclude that deepfakes are the most realistic and advanced form of synthetic media content. The quality of deepfake content is improving rapidly, due to ongoing advancements in artificial intelligence, particularly in deep learning (Westerlund, 2019).

The technology behind deepfakes has advanced dramatically in recent years. Modern deepfake systems utilize sophisticated generative adversarial networks (GANs) and deep learning techniques that can analyze existing footage or audio of a target individual and generate new, synthetic content that mimics their appearance and vocal patterns with remarkable accuracy (Wang et al., 2024). This technological leap has significantly lowered the barriers to creating convincing deepfakes, making these attacks accessible to a broader range of threat actors (Korshunov & Marcel, 2019).

Deepfake is increasingly exploited in various forms of social engineering, particularly within organizational contexts. These manipulations can be broadly categorized into two types: broadcast and narrowcast attacks (Bateman, 2020). Narrowcast deepfakes are tailored to deceive specific individuals through private communication channels. A critical organizational threat in this category is deepfake-enabled payment fraud. This attack method impersonates executives or trusted

partners, often via manipulated audio or video, to trick employees into authorizing illegitimate transactions or information. Payment fraud focuses on direct financial or information gains and represents a sophisticated evolution of traditional spear-phishing, now augmented by deepfakes (Bateman, 2020).

In contrast, broadcast deepfakes aim to deceive large audiences via public platforms and include tactics such as fabricating events to manipulate stock prices, deploying AI-generated personas or bots to sway investor sentiment, and initiating malicious bank runs by spreading false information (Bateman, 2020). While these methods primarily target public perception and stakeholder behavior, their influence on financial markets and organizational reputation can be profound (Westerlund, 2019). This study does not neglect the existence of these broadcast types of deepfake deception and the influence they have on organisations. However, it focuses on the narrowcasting type as a social engineering method, tailored to a specific employee.

The integration of deepfakes into social engineering attacks increases cybersecurity challenges by making it harder to differentiate between genuine and manipulated content (Kaushal, 2024). This development has significant implications for both individuals and organizations (Luca & Zervas, 2016). Organizations face more cybersecurity risks as new social engineering tactics continue to evolve (Wazid, 2023).

2.1.1 Organizational risks

Organizations face increasing risk from narrowcast deepfake attacks, highly targeted, AI-generated impersonations aimed at specific individuals within a company (Pedersen, 2025). Attackers often conduct extensive reconnaissance, collecting voice samples and behavioral cues from online materials such as earnings calls, conference presentations, and social media posts (Bateman, 2020). This data is then used to craft synthetic content that mimics executives with high accuracy, down to speech patterns, tone, and even real-time video likenesses. The attacker's goal is to gain access to critical systems, data, or financial resources. For example, in the Hong Kong case, attackers used a deepfake video to simulate a multi-person video call with what appeared to be actual coworkers and executives, successfully deceiving a finance employee into transferring $25 million (CNN, 2024). This demonstrates how narrowcast deepfakes can bypass traditional verification processes and exploit human vulnerabilities of employees, resulting in significant financial losses (Rancourt-Raymond & Smaili, 2023).

As deepfakes become harder to tell apart from real media, the risks for organizations grow. These include financial losses, damage to their reputation, loss of trust from stakeholders, and even manipulation of stock prices (Mustak et al., 2023). Deepfakes can also have serious personal consequences. For example, CEOs may experience stress and other psychological effects, while companies can suffer both financially and in how they are seen by the public (Rancourt-Raymond & Smaili, 2023).

Many organizations remain vulnerable to deepfake attacks due to their adoption of broad, vendor-centric security solutions that are not explicitly designed to detect or prevent synthetic media threats (Pedersen, 2025). Research indicates that deepfakes can bypass traditional security controls that organizations have implemented. Hence, human detection by employees remains a crucial part of organisational resilience against deepfake-based threats (Singh et al., 2025). While organizations as a whole face significant exposure to deepfake-enabled fraud, it is often individual employees who become the direct targets and unwitting enablers of such attacks. Understanding individual risks is necessary when dealing with narrow-scope deepfake-based payment fraud methods.

2.1.2 Individual risks

While organizational risks of deepfake-enabled payment fraud are significant, the threat landscape extends to individuals within organizations. Deepfake attacks in payment fraud scenarios often target specific employees, such as finance officers, executive assistants, or even C-suite executives, through highly personalized social engineering tactics (Kaushik et al., 2024). The personal impact of being deceived by a deepfake is not limited to financial repercussions. Individuals who authorize fraudulent payments under false pretences may experience significant psychological stress, guilt, and reputational damage within their organization (Rancourt-Raymond & Smaili, 2023). Repeated exposure to deepfake-enabled fraud attempts can erode an individual's confidence in digital communications, leading to decision paralysis or excessive caution in legitimate transactions (Matli et al., 2024).

As described by Krombholz et al. (2014), social engineers exploit human vulnerabilities to persuade individuals into performing malicious actions or disclosing sensitive information. Consequently, payment fraud through deepfake-based social engineering always relies on the manipulation of individuals, implicitly placing them at risk. A more in-depth analysis of social engineering and human vulnerabilities will be provided in the next chapter of the literature review.

Academic literature underscores the difficulty individuals face in detecting deepfakes, especially as the technology becomes more sophisticated (Guo et al., 2021; Liz-Lopez et al., 2024). Traditional cues, such as inconsistencies in tone or video artifacts, are increasingly absent in high-quality deepfakes, making it hard for untrained individuals to distinguish between genuine and manipulated content (Firc et al., 2023b). This challenge is compounded in high-pressure scenarios typical of payment fraud, where rapid decision-making is required.

A study by Aaron (2020) suggests that education and training may be sufficient to combat deepfake-based payment fraud at the individual level. However, before a clear employee resilience strategy can be developed, a deeper understanding is needed of the factors that make individuals more or less susceptible to this type of social engineering attack.

2.1.3 Deepfake detection: humans vs machine

Due to the rapid advancements in deepfake technology, identifying deepfake videos is becoming increasingly challenging. Consequently, the ability to distinguish real videos from deepfakes is growing ever more important as these videos become more convincing (Chakraborty & Naskar, 2024). Humans have traditionally relied on perceptual and contextual cues, such as unnatural facial movements, inconsistencies in audio, or contextual knowledge, to identify manipulated media (Guo et al., 2021; Firc et al., 2023a).

However, as deepfake technology advances, these cues are becoming increasingly subtle or even absent, making detection ever more challenging (Liz-Lopez et al., 2024). Empirical studies consistently show that untrained individuals struggle to accurately identify high-quality deepfakes, with detection rates often only slightly better than random guessing (Groh et al., 2021). This difficulty increases in high-pressure or stressful situations typical of payment fraud scenarios, where rapid decisions are required and cognitive resources are strained (Matli et al., 2024). While training and awareness programs can improve detection rates to some extent (Aaron, 2020; Groh et al., 2021), human performance remains limited by cognitive biases, workload, and the increasing sophistication of synthetic media (Somoray & Miller, 2024).

Hence, machine-based detection systems have rapidly advanced in response to the growing threat of deepfakes, leveraging artificial intelligence and deep learning to identify subtle artifacts and inconsistencies. These systems analyze digital fingerprints, pixel-level anomalies, and temporal inconsistencies in audio-visual content, allowing for the detection of manipulated media at a scale and speed unattainable by human reviewers (Kietzmann et al., 2020). Benchmark studies, such as

those conducted during the ‘DeepFake Detection Challenge’, demonstrate that state-of-the-art algorithms can achieve high accuracy rates in controlled environments (Dolhansky et al., 2020). However, their effectiveness can diminish when confronted with novel or highly sophisticated deepfakes or when adversarial techniques are used to bypass detection (Verdoliva, 2020).

Furthermore, machine learning models require constant retraining and updating to keep pace with evolving deepfake generation methods (Kaushal, 2024). AI-generated voice deepfakes can fool voice recognition systems, and AI-manipulated images can deceive facial recognition software used for authentication purposes (Kassis & Hengartner, 2023; Bodepudi & Reddy, 2020). Simultaneously, for most large organizations, scaling deepfake models is (too) expensive, whereas small companies cannot implement them completely (Patel et al., 2023).o

Despite these limitations, machine detection remains a critical component of organizational defense, as it can process vast amounts of content rapidly and flag suspicious media for further human review (Westerlund, 2019). As the literature suggests, the most robust approach combines automated detection with human oversight (Singh et al., 2025). While significant academic research has focused on technological tools for detecting deepfakes, there is limited understanding of human-based differences in deepfake detection (Xu et al., 2022), which creates a research gap in an important line of defence against deepfake fraud. Social engineering theory could provide more insights into human vulnerabilities that are exploited in deepfake-based payment fraud methods.

2.2 Defining deepfake-based social engineering

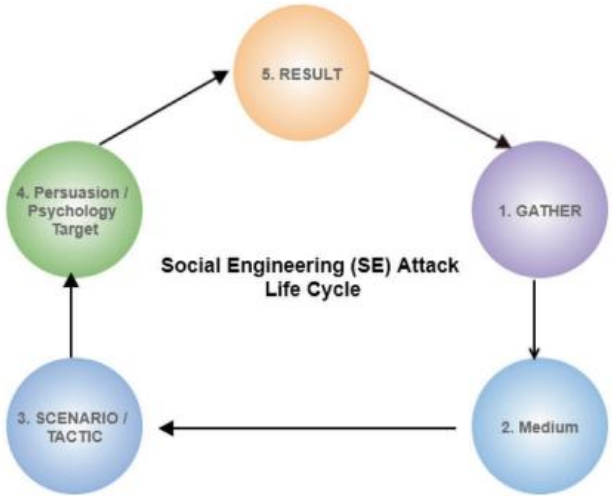
Social engineering involves manipulating individuals, often through psychological persuasion, to disclose confidential information or grant access to secure systems (Mitnick & Simon, 2002). Despite advancements in technical security, the human factor remains a key vulnerability due to its susceptibility to manipulation (Mouton et al., 2016). Attackers exploit this weakness by using persuasive techniques to prompt individuals to share sensitive data or carry out harmful actions. Such attacks frequently combine psychological, social, and technical strategies, applied at various stages of the intrusion process to achieve their objectives (Krombholz et al., 2015).

Yasin et al. (2021) introduce the Social Engineering Attack Life Cycle (SEALC), which conceptualizes social engineering attacks as a five-step process. This framework, illustrated in Figure 1, outlines each phase of an attack and provides a structured understanding of how such attacks typically unfold. While the model is relatively general, it serves as a valuable foundation for

grasping the fundamental dynamics of social engineering. In the remainder of this chapter, we delve deeper into the SEALC and explore its relevance to deepfake-based payment fraud.

Figure 1

Social engineering attack life cycle



Note. Adapted from Yasin et al. (2021).

1. GATHER: The attacker collects detailed information about the target and creates a convincing story or scenario to use in the attack.
2. MEDIUM: The attacker initiates contact using a chosen communication channel, often using tactics like number spoofing or malicious software to increase credibility and impact.
3. SCENARIO: The attacker presents the fabricated problem or situation to the victim, aiming to make the story believable using the data gathered earlier.
4. PERSUASION: Psychological manipulation is used to exploit the victim's emotions or weaknesses, through fear, urgency, or rewards, to prompt a specific action.
5. RESULT: The attacker achieves their goal, such as obtaining confidential information, financial gain, or system access, often leading to further exploitation (Yasin et al., 2021).

The use of deepfakes as a medium in social engineering attacks presents a significant challenge to cybersecurity by making it increasingly difficult to recognize manipulated media (Dsouza et al., 2024). These synthetic videos and audio clips have reached a level of sophistication where they can convincingly imitate real human appearances and voices (Van der Sloot & Wagensveld, 2022). As a result, attackers can create compelling scenarios, such as fake video calls or voice messages, that

appear to come from trusted sources. These fabrications deceive victims into sharing confidential information or performing actions they would not normally consider (Mirsky & Lee, 2021).

In the fourth phase of the SEALC, persuasion, attackers exploit human vulnerabilities by manipulating psychological constructs that influence behavior and decision-making. These constructs make individuals more susceptible to deception. A deeper understanding of these human vulnerabilities and the mechanisms through which they are exploited is important. Insight into potential vulnerabilities will be necessary to address these vulnerabilities and reduce people's susceptibility to fraud.

2.2.1 Human vulnerabilities

Human vulnerabilities are the inherent weaknesses or susceptibilities in human behavior and decision-making that cybercriminals exploit through social engineering (Siddiqi, 2022). Unlike technical vulnerabilities, these are rooted in psychology and social dynamics. Referring back to the SEALC, persuasion or psychological weakness focuses on exploiting human vulnerabilities, making people susceptible to manipulation. From an information systems research perspective, understanding why individuals fall for these attacks requires a detailed investigation into human cognition, psychological triggers, and behavioral terms exploited by attackers. One area of study that helps in uncovering a better method of combating social engineering is the field of psychology, specifically social psychology (Scheeres, 2008). First, human vulnerability within social engineering is discussed; then, the scope is narrowed to deepfake-based financial fraud. The main reason is the inconclusiveness and lack of human-centric deepfake detection literature.

Gragg (2003) made a significant contribution to the field by identifying seven psychological triggers commonly exploited in social engineering attacks: strong affect, cognitive overload, reciprocation, deceptive relationships, diffusion of responsibility, authority, and integrity/consistency. These triggers closely mirror behaviors relevant to the information systems context, such as following internal IT directives or responding to urgent-looking emails without scrutiny. When an attacker effectively leverages these psychological mechanisms, individuals often comply and reveal sensitive information with minimal resistance. Because social engineering fundamentally relies on persuasion and deception, it is essential to investigate how individuals can identify and resist such manipulative strategies.

Expanding on these insights, Montañez et al. (2020) developed a framework that highlights key psychological factors affecting vulnerability to social engineering. Their research identifies high

cognitive workload, increased stress, diminished attentional vigilance, lack of domain-specific knowledge, and limited prior experience as factors that significantly increase susceptibility to social engineering attacks. As stated by Montanez et al. (2020), while personality traits also influence susceptibility, the literature on this relationship remains inconclusive about the differences in how the Big Five personality traits influence susceptibility, leaving the question of whether these psychological and cognitive factors also influence deepfake detection.

Another significant study on human vulnerabilities in social engineering by Wang et al. (2021) highlights that "individuals' personality traits significantly contribute to their susceptibility to social engineering exploits such as influence, manipulation, and deception". Social engineers exploit these personality traits as psychological vulnerabilities, using deception and persuasive language to manipulate their targets. This makes personality traits a relevant and emerging area of interest in the context of deepfake detection, where tailored manipulation strategies are increasingly employed.

In addition to personality, vulnerabilities in human cognition, particularly cognitive workload, are consistently identified in the literature as critical factors contributing to the effectiveness of social engineering attacks (Gragg, 2003; Montañez et al., 2020; Wang et al., 2021; Dsouza, 2024). Therefore, research into human vulnerabilities in deepfake detection should incorporate a focused examination of both cognitive workload and personality traits. Recognizing these psychological and cognitive factors is crucial for developing targeted interventions and training programs aimed at reducing susceptibility to deepfake-based social engineering.

Shifting the focus from traditional social engineering tactics to those involving deepfakes, a recent study by AlSobeh et al. (2024) offers valuable insight into the psychological vulnerabilities exploited by this emerging threat. Their exploratory study, conducted among students from two Midwestern universities, examined how psychological and cognitive traits and digital literacy influence individuals' ability to detect deepfakes. The findings suggest that individuals with stronger analytical thinking skills and higher levels of digital literacy were more successful in identifying manipulated media. However, the study's generalizability is limited due to its small sample size and the exclusive use of student participants from the USA. Moreover, the study addressed a broader scope, also exploring the psychological effects of exposure to deepfakes, rather than focusing specifically on detection mechanisms. This highlights a gap in the literature concerning the specific role of psychological and cognitive factors in accurately identifying deepfakes.

Dsouza et al. (2024) argue that the integration of deepfakes into conventional social engineering methods significantly increases the sophistication of such attacks. By leveraging emotional, cognitive, and behavioral cues, attackers can enhance the effectiveness of deception. A central theme in the work of Dsouza et al. (2024) is the erosion of trust as a human vulnerability. As manipulated video and audio content become increasingly realistic, individuals find it more difficult to distinguish between authentic and fabricated information. This blurring of authenticity compromises the traditional markers used to assess credibility, thereby increasing susceptibility to manipulation (Holiday, 2018). Consequently, trust and credibility emerge as attack mechanisms exploited in deepfake-based social engineering attacks. However, despite the relevance of trust and credibility as attack mechanisms in deepfake-based social engineering attacks, this study will not focus on them. The intricate nature of trust and its role in deception, though essential, would require a more in-depth exploration that exceeds the scope and resource limitations of the current study. By narrowing the focus, this study aims to offer targeted insights that can directly inform interventions and training programs for improving detection accuracy.

As deepfakes increasingly become embedded in sophisticated social engineering attacks, the psychological manipulation of individuals emerges as a core strategy used by attackers (Dsouza et al., 2024). The interplay between human vulnerabilities, such as stress, cognitive overload, and personality traits, and the realism of synthetic media greatly enhances the persuasiveness of these attacks (Wang et al.,2021; Montanez et al., 2020; Gragg, 2003). Although organizations can deploy technological detection solutions, human judgment remains a critical line of defense (Singh et al., 2025). To effectively equip individuals to detect and resist deepfake deception, it is necessary to understand the underlying psychological and cognitive mechanisms that influence their decision-making processes. The following sections offer a deeper analysis of the psychological and cognitive factors involved.

2.2.2 Key effect mechanisms

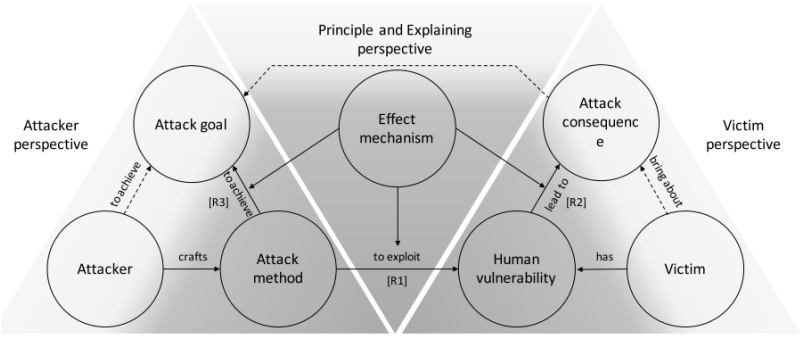
Further elaborating on social engineering attacks and their exploitation of human vulnerabilities, Wang et al. (2021) present a clear and structured conceptual framework that explains social engineering attacks through the effect mechanisms that enable them. Analysing this framework provides valuable insight into how specific cognitive and psychological factors are targeted and manipulated. The model illustrates the dynamic interaction between the attacker and the victim, mediated by mechanisms that facilitate the attack. From the attacker’s perspective, a specific attack

goal is pursued by crafting a method designed to exploit human vulnerabilities via specific effect mechanisms. These mechanisms function as the core enablers of the attack.

On the victim's side, these vulnerabilities lead to specific consequences, ultimately affecting the victim. The model highlights a cause-and-effect chain: from the attacker's goal, via method and mechanism, to the impact on the human target. This dual-perspective framework, attacker and victim, helps explain not only how social engineering works but also when psychological and cognitive factors are most susceptible to manipulation. The conceptual framework is illustrated in Figure 2.

Figure 2

Conceptual framework for social engineering attacks



Note. Adapted from Wang et al. 2021.

Multiple effect mechanisms often interact to form a more sophisticated and convincing attack method. This highlights how attackers strategically design attack scenarios that integrate various techniques, psychological triggers, and human vulnerabilities in a coordinated manner to achieve their intended goals (Wang et al., 2021). In the case of deepfake-based payment fraud, for instance, mechanisms such as obedience to authority or trust in familiar individuals play a crucial role and cannot be overlooked (Pedersen et al., 2025; Dsouza et al., 2024).

This study, however, focuses on cognitive and psychological effect mechanisms that make individuals susceptible to social engineering, particularly in the context of deepfake-enabled fraud. Several foundational theories from cognitive and social psychology help explain how attackers exploit human vulnerabilities to influence behavior.

The Elaboration Likelihood Model (ELM) describes two persuasion routes: a central route involving careful analysis, and a peripheral route relying on superficial cues like authority or

urgency. Attackers typically steer victims toward the peripheral route, reducing critical thinking and increasing compliance (Pertier, 2006). The Cognitive Response Model complements this by highlighting how distraction, emotional arousal, or overload weakens a victim's ability to generate counterarguments (Petty, 1977).

Another important effect mechanism is time pressure and cognitive overload. Time pressure occurs when individuals must process a large amount of information within a limited time. Time pressure within social engineering attacks has been proven to reduce detection ability. Within phishing attacks, numerous studies have demonstrated that time pressure, as evidenced by urgent calls, significantly reduces employees' detection abilities (Butavicius et al., 2022). It can also trigger emotional responses like stress, further impairing cognitive abilities (Chowdhury, 2020). The danger of cognitive overload happens when complex or excessive information overwhelms cognitive capacity, making it difficult to process or evaluate content critically (Gragg,2003). Time pressure will be adapted as a contextual variable in this study, which is added to the scenario.

A comprehensive analysis of these theories and their implications for deepfake detection is provided in Chapter 2.3.

2.2.3 Psychological factors in deepfake detection

As deepfake technology continues to evolve, the detection of synthetic media has become an increasingly complex challenge. While prior sections have established that deepfakes can bypass traditional security controls and exploit human vulnerabilities, recent academic research highlights the critical role of individual psychological factors in determining susceptibility to deepfake deception (Dsouza, 2024; AlSobeh, 2024). In particular, studying human vulnerabilities identified personality traits and stress (Montanez et al., 2020; Wang et al., 2021) as key psychological constructs that help explain why some individuals are more susceptible to social engineering attacks than others.

Personality traits

Individual personality traits play a significant role in influencing susceptibility to social engineering tactics such as manipulation, persuasion, and deception, with phishing being the most widely researched form of social engineering (Uebelacker et al., 2014). Although no studies have directly examined the relationship between personality traits and deepfake detection accuracy, insights from phishing research suggest meaningful associations between personality and vulnerability to digital deception. Personality traits, shaped by both genetic and environmental influences, are relatively

stable across time and consistently affect behavior across various contexts. The Five-Factor Model (FFM) provides a widely accepted framework for categorizing personality into five dimensions: extraversion, conscientiousness, agreeableness, openness to experience, and neuroticism (McCrae & John, 1992).

Extraversion is defined by traits such as sociability, assertiveness, activity, and a tendency to seek excitement and positive emotions. Extraverts are typically outgoing, energetic, and enjoy social interaction (McCrae & John, 1992). However, these very characteristics can make them more susceptible to phishing attacks. Uebelacker et al. (2014) found that extroverted individuals are more likely to comply with requests, which increases their vulnerability. Empirical studies reinforce this pattern: both Cusack and Adedokun (2018) and Darwish et al. (2012) reported that individuals high in extraversion are more likely to trust strangers and respond to authority, making them easier targets for phishing attempts. Moreover, Lawson et al. (2018) demonstrated that extraversion is negatively associated with phishing detection accuracy. In sum, the social responsiveness and trusting nature of extraverts heighten their risk in phishing contexts.

Conscientiousness, in contrast, is generally a protective trait. It encompasses attributes such as self-discipline, orderliness, responsibility, and careful planning (McCrae & John, 1992). Conscientious individuals tend to adhere to rules and protocols, which often translates into secure online behaviors and better phishing detection outcomes (Uebelacker et al., 2014). Studies by Halevi et al. (2016) and Lawson et al. (2018) support this, indicating that high conscientiousness correlates with reduced phishing susceptibility and more secure online behavior. However, Halevi et al. (2015) also caution that highly conscientious individuals who underestimate risks may become overconfident or fall into habitual behavior patterns, which could make them vulnerable in specific scenarios. Time pressure can also break their reliance on structured processes. The stress impairs their ability to engage in the careful analysis they typically rely upon (Vollrath, 2001). Overall, conscientiousness acts as a buffer against social engineering, but conscientiousness does not make people foolproof. Without sufficient training, conscientiousness could lead to a higher susceptibility.

Agreeableness, reflecting trust, empathy, modesty, and a cooperative nature (McCrae & John, 1992), presents a more consistent risk factor in the phishing context. Agreeable individuals often exhibit a strong desire to help and avoid conflict, making them prone to manipulation. Cho et al. (2016) found a clear association between high agreeableness and increased phishing susceptibility. This vulnerability stems from their trusting nature and low tendency to question requests, especially those framed by authority figures or appearing altruistic. Cho et al. (2016) further reported that

agreeable individuals tend to perceive phishing messages as less risky, reducing their likelihood of recognizing malicious intent. This implies that agreeableness should make people more susceptible to attacks.

Openness to experience includes traits such as intellectual curiosity, creativity, appreciation for art, and willingness to entertain new ideas (McCrae & John, 1992). The relationship between openness and phishing susceptibility is more nuanced. On the one hand, high openness can foster skepticism and analytical thinking, which help detect suspicious messages. Lawson et al. (2018) and Pattinson et al. (2012) found that individuals high in openness are better at distinguishing legitimate emails from phishing attempts, possibly due to greater self-efficacy in managing digital environments. On the other hand, Halevi et al. (2013) noted that openness is also linked to lower privacy settings on social media and a greater tendency to explore unfamiliar platforms, increasing exposure to social engineering threats. Therefore, while openness can enhance detection accuracy, it also entails behavioral risks that may increase overall vulnerability.

Finally, neuroticism is characterized by emotional instability, anxiety, impulsiveness, and vulnerability to stress (McCrae & John, 1992). Individuals high in neuroticism often experience greater fear and worry, which can both increase and decrease susceptibility to phishing. Byrne et al. (2015) argue that emotional manipulation techniques, such as exploiting cognitive dissonance, are particularly effective against neurotic individuals. Halevi et al. (2013) found that high neuroticism is associated with poor cybersecurity behaviors and greater responsiveness to emotionally charged phishing attacks, such as scams promising rewards or inducing urgency. However, Cho et al. (2016) suggest that the heightened anxiety associated with neuroticism can also lead to increased risk perception and reduced trust, potentially making individuals more cautious. Thus, although individuals who score high on neuroticism may exhibit increased vigilance due to anxiety, they are also more vulnerable to fear-based manipulation.

In summary, the Five-Factor Model offers valuable insights into individual differences in phishing susceptibility, with potential implications for deepfake detection as well. Extraversion, agreeableness, and neuroticism generally increase vulnerability, while conscientiousness provides a protective buffer, and openness has mixed effects depending on contextual factors. As deepfakes increasingly mimic the persuasive strategies found in phishing, particularly those exploiting trust, urgency, and authority, the role of personality traits in shaping detection ability needs further investigation. By linking personality traits to detection accuracy, this study aims to uncover a

correlation between individual differences and deepfake detection. This connection is particularly relevant in understanding who is most at risk in deepfake-enabled social engineering.

Stress resilience

Stress has been a key variable of interest in social engineering research. Montañez et al. (2020) identified it as the second most important factor increasing susceptibility to social engineering attacks. Similarly, AlSobeh (2025) found a correlation between higher stress levels and lower deepfake detection accuracy, underscoring the predictive power of stress in determining detection performance. This raises the question of whether individuals who are more resilient to stress are also more effective at detecting deepfakes, particularly in high-stress situations.

The Sense of Coherence (SOC) framework assesses an individual's ability to perceive life as structured (comprehensible), manageable (solvable), and meaningful (purposeful) (Antonovsky, 1993). Gee et al. (2018) emphasize its role in fostering resilience to stress and supporting long-term well-being. A strong SOC has been shown to help regulate stress levels and influence how individuals cope with stressful situations on the workfloor (Jenny et al., 2017). Employees with a higher SOC may therefore be less vulnerable to deepfake-based social engineering attacks, particularly when stressors such as time pressure are introduced into the scenario.

2.2.4 Cognitive factors in deepfake detection

Cognitive workload is a critical aspect of human cognition that influences susceptibility to social engineering attacks (Gragg, 2003; Montañez et al., 2020; Wang et al., 2021; Dsouza, 2024). It refers to the mental effort required to perform a task and is shaped by both the complexity of the task and an individual's cognitive capacity. Depending on task demands, individuals may be able to manage multiple activities simultaneously without a decline in performance, indicating a manageable workload, or they may become cognitively overloaded, resulting in reduced performance (Kosch et al., 2023). Research into cognitive workload is ongoing, with the development of more refined tools, theoretical frameworks, and practical guidelines expected in the coming years (Duran et al., 2022). However, despite its relevance, research that focuses explicitly on the role of cognitive workload in deepfake detection remains limited and underexplored.

Ask et al. (2023) found that individuals with higher cognitive flexibility are more accurate in evaluating their ability to detect deepfakes. This suggests that the capacity to shift between strategies or mental frameworks facilitates the navigation of deceptive content. However, since cognitive flexibility is considered a relatively stable, trait-like characteristic, it cannot fully account

for variations in performance across different situations. This underscores the importance of examining dynamic, situational factors, such as cognitive workload, which fluctuate with task demands and environmental pressures and may have a more immediate impact on an individual's susceptibility to deception (Kosch et al., 2023).

Similarly, Ahmed et al. (2021) found that higher cognitive ability correlates with greater accuracy in deepfake detection. However, this finding is more reflective of general cognitive capacity rather than performance within a specific task context.

One key situational factor influencing cognitive workload is time pressure. While task difficulty also contributes to cognitive load, time pressure intensifies it by requiring individuals to allocate mental resources within a restricted timeframe. This often impairs information processing efficiency and reduces overall performance (Galy et al., 2017).

In this literature review, cognitive workload, personality traits, and sense of coherence are identified as potentially relevant variables influencing human deepfake detection. In the following chapter, information processing theory will be explored, with a focus on the mechanisms of the Elaboration Likelihood Model, as well as the roles of time pressure and cognitive overload. Understanding these theoretical constructs is essential for formulating the research hypotheses.

2.3 Information processing theory

The Elaboration Likelihood Model (ELM) helps explain how people process information and assess credibility. Cacioppo et al. (1986) developed the ELM to explore a wide range of variables that influence the likelihood of elaboration and, consequently, the route to persuasion. According to the model, people process persuasive information through two main routes: the central route, where individuals critically analyze content when they are motivated and able, leading to more durable attitude changes; and the peripheral route, where individuals rely on superficial cues when motivation or ability is low, resulting in more temporary persuasion.

In the context of deepfake video detection, Goh (2023) illustrates how individuals assess such videos through the lens of the ELM. Two primary strategies are identified: media-based identification strategies (e.g., spotting visual anomalies), which align with peripheral processing, and knowledge-based strategies (e.g., verifying sources), which align with central processing. The study explains that combining both strategies leads to the most effective detection outcomes. However, optimal detection practices may vary depending on contextual factors such as the video's quality and duration, as well as individual differences in cognitive processing. For instance, when

short video clips are presented, people tend to rely more on the peripheral route, as limited context hinders deeper evaluation via the central route. In such cases, individuals are more dependent on superficial cues, as understanding the message itself becomes less relevant for assessing authenticity.

In social engineering research, the usage of time pressure as an effect mechanism has been widely recognized (Butavicius et al, 2022; Chowdhury, 2020; Gragg, 2003; Galy et al., 2017). The use of time pressure as a situational factor or effect mechanism is therefore likely in deepfake-based payment fraud. This means that a media-based identification strategy will be necessary to identify these deepfake frauds. The five media-based identification strategies consist of graphical anomalies, behavioral anomalies, production quality, voice inflection, and sound quality (Goh, 2023).

In light of this, a new variable will be crucial: ‘experience in deepfake detection’. Such experience provides individuals with a trained ability to recognize graphical anomalies or apply other media-based strategies more effectively (Tomlinson, 2022). Therefore, experience in deepfake detection is introduced as an independent variable, which is expected to enhance detection accuracy under conditions of time pressure, in contrast to the study *Fooled Twice* by Kobis et al. (2021), which found that IT and cybersecurity professionals did not outperform non-IT professionals in detecting deepfakes. Detection accuracy might be higher for people with more detection experience. Working in IT and cybersecurity alone does not give you an advantage, unless detecting deepfakes is part of your job.

2.4 Conceptual framework

The rapid advancement of deepfake technology has introduced significant new risks for organizations, particularly in the realm of social engineering and payment fraud. While technical solutions for deepfake detection continue to evolve, the human element remains a critical line of defence, especially as attackers increasingly target employees with highly realistic synthetic media. Understanding which cognitive and psychological factors influence employees’ ability to detect deepfakes, especially under time pressure, is essential for developing effective organizational countermeasures.

The relevance and influence of each psychological and cognitive factor on deepfake detection have been discussed in the previous chapter. This section aims to summarize the variables included in the study rather than revisit their theoretical foundations. The research is grounded in the ELM and

incorporates psychological and cognitive constructs that contribute to human vulnerability in the context of social engineering.

The ELM explains how people assess credibility via two routes: central (analytical) and peripheral (superficial). In deepfake detection, individuals often rely on peripheral, media-based cues, especially under time pressure. Experience in detecting deepfakes enhances the use of these cues, improving detection accuracy in high-pressure scenarios.

H1: Prior deepfake detection experience positively affects deepfake detection accuracy.

The previous chapters explained how attackers exploit human vulnerabilities. Personality traits and stress are key psychological factors influencing susceptibility to deepfake deception. Traits like extraversion, agreeableness, and neuroticism tend to increase vulnerability, while conscientiousness is generally protective, and openness could show moderately decreased effects. Additionally, individuals with higher stress resilience, measured through the Sense of Coherence, may be better equipped to detect deepfakes under pressure. These factors help explain individual differences in detection performance.

H2: Personality Traits

H2a: Extraversion negatively affects deepfake detection accuracy.
H2b: Agreeableness negatively affects deepfake detection accuracy.
H2c: Neuroticism negatively affects deepfake detection accuracy.
H2d: Conscientiousness positively affects deepfake detection accuracy.
H2e: Openness to experience positively affects deepfake detection accuracy.

H3: Sense of coherence positively affects deepfake detection accuracy.

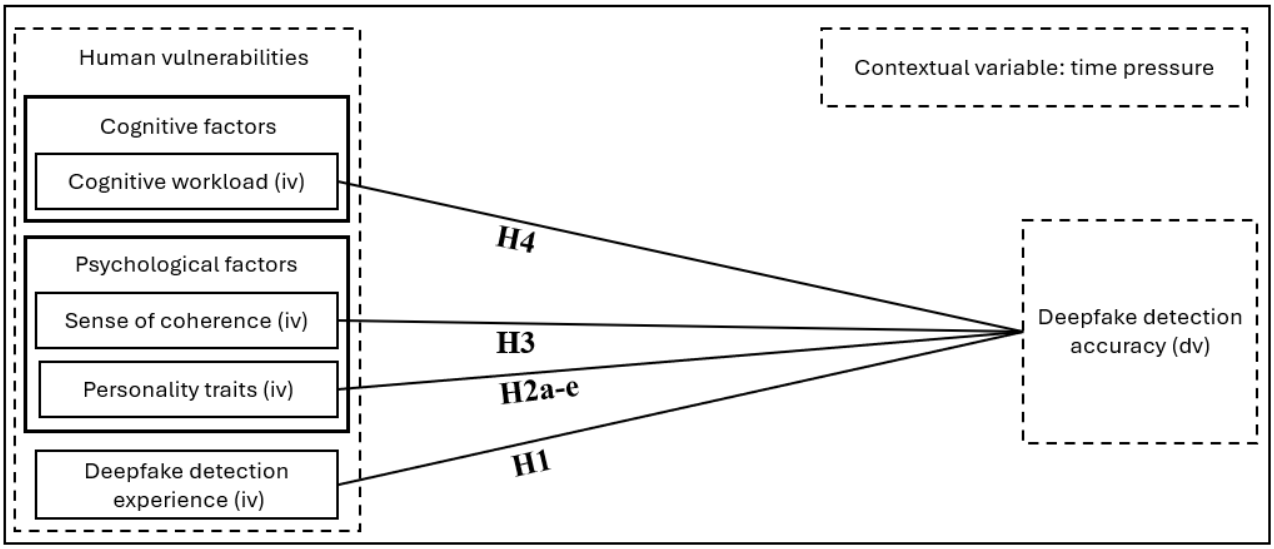
Cognitive workload refers to the mental effort needed to complete a task and influences susceptibility to deception. High workload, especially under time pressure, can overwhelm cognitive capacity and reduce deepfake detection accuracy. As a situational factor, time pressure amplifies workload, impairing information processing and decision-making.

H4: Cognitive workload negatively affects deepfake detection accuracy.

Figure 3 presents the conceptual model outlining the human factors influencing deepfake detection accuracy. This model illustrates how various human vulnerabilities and contextual factors interact to affect an individual's ability to detect deepfakes.

Figure 3.

Conceptual framework



Note. This conceptual framework illustrates the hypothesized relationships between individual psychological and cognitive vulnerabilities and the ability to detect deepfakes.

The conceptual framework groups independent variables under 'Human vulnerabilities,' subdivided into cognitive and psychological factors, as well as prior detection experience. This structure reflects the multidimensional nature of susceptibility to deepfake deception while maintaining analytical tractability for empirical research. However, the linear relationships implied by these hypotheses may not fully capture the complexity of human decision-making under pressure, where multiple factors could interact in non-additive ways to influence detection performance. This complexity necessitates an additional analysis. A configurational approach is therefore employed to identify how different combinations of variables interact to produce varying levels of detection accuracy (Furnari et al., 2021).

Propositions for configurational effects

Configurational theory emphasizes the principle of equifinality, which holds that multiple, distinct combinations of causal conditions can produce the same outcome (Woodside, 2014). In the context of deepfake detection, varying combinations of psychological, cognitive, and experiential variables may lead to either higher or lower detection accuracy, depending on how these factors interact. Prior research on personality traits, for instance, has produced mixed and sometimes contradictory results (Frauenstein & Flowerday, 2020). The trait of openness, for example, has been linked to both analytical thinking and skepticism, attributes that could enhance detection accuracy.

Conversely, it has also been associated with riskier online behavior, such as weaker privacy settings and a greater inclination to explore unfamiliar platforms, thereby increasing exposure to social engineering threats (Halevi et al., 2013).

A configurational approach is therefore essential for identifying such nuanced and potentially non-linear patterns. This perspective also introduces the concept of causal asymmetry, which suggests that the presence or absence of a particular outcome depends on the specific combination of conditions, rather than any single factor acting independently (Woodside, 2014). Based on this reasoning, two propositions are formulated:

Proposition 1. Different configurations of detection experience, personality traits, sense of coherence, and cognitive workload can lead to a high deepfake detection experience.

Proposition 2. Different configurations of detection experience, personality traits, sense of coherence, and cognitive workload can lead to low deepfake detection experience.

The propositions enrich this research by moving beyond the siloed and linear nature of traditional hypotheses.

3 Research methodology

The study adopts a quantitative explanatory design to investigate the relationships between psychological and cognitive factors, specifically personality traits, cognitive workload, sense of coherence, detection experience, and the accuracy of deepfake detection under time constraints. This approach seeks to assess the strength and direction of associations between variables, whilst also investigating what combinations of variables heighten detection accuracy.

The current design focuses on correlation, which is appropriate because key constructs such as personality and sense of coherence represent stable individual differences that cannot be ethically or practically manipulated (Price, 2015). Rather than attempting to manipulate these inherent traits, the study adopts a naturalistic approach that examines how existing individual differences relate to task performance under standardized conditions. This approach allows for the investigation of naturally occurring relationships while maintaining control over task presentation and measurement procedures. The variables are analyzed about participants' performance on a deepfake detection task under the contextual condition of time pressure. Time pressure is not manipulated. All participants are exposed to the same time constraints. This study does not examine the moderating effect of time pressure but rather explains the potential relationships between psychological and cognitive factors and deepfake detection accuracy, whilst being under time pressure.

To increase both reliability and validity, standardized questionnaires are used to measure the variables of interest (Greavu-Şerban et al., 2025). Cognitive workload is assessed after the task using the NASA Task Load Index (NASA-TLX). Sense of coherence is measured beforehand using the SOC-13 scale, while personality traits are assessed prior to the detection test using the BFI-10 questionnaire. Additionally, participants report their prior experience with deepfake detection.

3.1 Justification for method selection

The selection of appropriate research methods is a foundational element in ensuring the validity, reliability, and overall quality of academic research (Creswell & Creswell, 2018). In this study, the chosen methods were selected through a careful alignment with the research objectives, the nature of the research question, and the characteristics of the available data (Bryman, 2016).

The primary objective of this research is to identify what cognitive and psychological factors impact employees' deepfake detection accuracy under time pressure. According to Saunders, Lewis, and Thornhill (2019), the research method must be consistent with the research aims and the type of

knowledge sought. Given that this study seeks to explain, a quantitative approach was deemed most appropriate. This choice enables a comprehensive examination of different psychological and cognitive factors, ensuring both depth and breadth in data collection and analysis.

As noted by Flick (2018), the method must be suitable for the type of data and the phenomenon under investigation. For example, qualitative methods such as thematic analysis are particularly effective for identifying patterns and meanings within textual data (Braun & Clarke, 2006), while quantitative methods such as regression analysis are well-suited for examining relationships between variables (Field, 2018). To complement the quantitative explanatory approach, this study employs fsQCA as a configurational methodology that transcends the traditional qualitative-quantitative divide by examining how multiple conditions combine to produce outcomes rather than focusing on net effects of individual variables (Ragin, 2008).

The selection process was further informed by established methodological literature. For instance, the use of quantitative explanatory design is widely supported in similar research contexts (AlSobeh et al., 2024; Ahmed & Chua, 2023; Sütterlin et al., 2022). This method allows to examine the relationships among variables in a structured and systematic manner. Creswell & Creswell (2018) describe explanatory research as a process where the researcher begins with a theory or hypothesis, collects quantitative data, and then uses statistical procedures to test the hypothesized relationships. This approach is particularly effective for identifying patterns, associations, and potential causal links between variables.

Alternative methods, such as interviews, were considered. However, these were deemed less suitable due to several key reasons. First, interviews and other qualitative approaches, while valuable for exploring subjective experiences or generating in-depth insights, are less effective when the research objective is to quantify relationships and test theoretical propositions across a larger sample (Creswell & Creswell, 2018; Bryman, 2016). Qualitative methods are inherently limited in their ability to provide generalizable findings or to statistically assess the strength and direction of associations between variables (Field, 2018).

Moreover, the constructs of interest in this study, such as personality traits, sense of coherence, and cognitive workload, are well-established in the literature and are most reliably and validly measured using standardized quantitative instruments (Greavu-Şerban et al., 2025). As recommended by Yin (2018), the selection of methods should always be justified through a comparison of strengths and limitations about the research aims.

3.2 Measurement instruments

The dataset used in this study was collected through a structured online survey that combined standardized psychological instruments with a deepfake detection task. The aim is to explain how cognitive and psychological variables relate to deepfake detection accuracy. Each variable was measured using validated tools to ensure reliability and construct validity. The dependent variable is deepfake detection accuracy, derived from a detection task based on the methodology of Köbis et al. (2021). The resulting accuracy score, expressed as a percentage, reflects their detection performance. An overview of study variables, corresponding measurement instruments, data types, and scoring scales is illustrated in table 1.

The psychological variables assessed include personality traits and sense of coherence, both measured using validated instruments. Personality traits are measured using a brief version of the Big Five Personality Inventory, called the BFI-10, which has demonstrated a test–retest reliability of approximately 0.75 (Rammstedt & John, 2006). The BFI-10 was selected over the longer BFI-44 to reduce survey length and participant burden (Appendix A). Sense of coherence is measured using the SOC-13, which has shown consistently high reliability across various countries, with reported Cronbach's alpha values ranging from 0.81 to 0.93 (Antonovsky, 1993) (Appendix B). These instruments were chosen for their strong psychometric properties and efficient formats.

In the post-test phase, participants complete the NASA Task Load Index (NASA-TLX) to assess cognitive workload across six dimensions: mental demand, physical demand, temporal demand, performance, effort, and frustration. The NASA-TLX has demonstrated strong reliability, with intraclass correlation coefficients (ICCs) ranging from 0.71 to 0.81 (Devos et al., 2020) (Appendix C).

Over time, the NASA-TLX method has become widely adopted across various socio-technological domains. It is now the most commonly used tool in human-computer interaction studies to measure cognitive workload (Kosch et al., 2023). The NASA-TLX serves as a subjective measure of cognitive workload (Kosch et al., 2023). Objective measures for cognitive workload, such as Doppler transcranial blood flow measurement or electroencephalogram (EEG), would be well suited (Khan et al., 2023). However, for this study the NASA-TLX was chosen due to practical considerations. Replicating the study with objective measurement techniques could provide valuable new insights.

Following the pre-test, participants complete a deepfake detection task, adapted from the experimental design of Köbis et al. (2021), using their validated video database. Participants are shown ten short video clips, each approximately 10 seconds in length, five real and five deepfakes. To induce time pressure, participants are allowed to view each video only once, without the option to replay it. This constraint creates a sense of urgency and limits reflective decision-making. Prior to the test, participants receive a briefing (outlined in Appendix D). Unlike in the original Köbis et al. (2021) study, participants are not informed about the ratio of real to fake videos. This design choice aims to reduce potential bias caused by expectancy effects related to video proportions. After viewing each video, participants must classify it as either a deepfake or a real video (Yes/No).

Table 1. Overview of study variables, corresponding measurement instruments, data types, and scoring scales.

Variable	Instrument	Data type	Scale
Deepfake detection accuracy (DV)	Detection experiment by Köbis et al. (2021)	Ratio	Accuracy score (0–100%)
Cognitive workload (IV)	NASA-TLX (Hart, 2006)	Interval	Weighted score (1-20)
Sense of coherence (IV)	SOC-13 (Antonovsky, 1993)	Interval	Average score (1-7)
Extraversion (IV)	BFI-10 (Rammstedt & John, 2006)	Interval	Average score per item (1-5)
Agreeableness (IV)	BFI-10 (Rammstedt & John, 2006)	Interval	Average score per item (1-5)
Conscientiousness (IV)	BFI-10 (Rammstedt & John, 2006)	Interval	Average score per item (1-5)
Neuroticism (IV)	BFI-10 (Rammstedt & John, 2006)	Interval	Average score per item (1-5)
Openness (IV)	BFI-10 (Rammstedt & John, 2006)	Interval	Average score per item (1-5)
Deepfake detection experience (IV)	None	Interval	7-Point Likert scale

3.3 Sampling strategy and participants

The minimum sample size ranges from 52 to 109 participants, calculated using a prior G*Power analysis (Faul et al., 2009). The effect size ranges from medium to large ($f = 0.25$ to 0.40), with a statistical power of 0.80 and an alpha level of 0.05 (Raywood-Burke, 2023). As described in the

data section, eight predictors, serving as independent variables, are used to predict detection accuracy; in total, 89 participants filled in the survey.

As previous findings suggest, age, gender, education, and profession do not play a significant role in influencing detection abilities (Ahmed, 2021; Tahir et al., 2021). The main inclusion criteria were employment within a computer-based working environment and a sufficient level of English proficiency. All participants were recruited via social media and personal contacts within the researcher's network. They were based in various European countries, with the majority residing in the Netherlands.

3.4 Ethical considerations

The chosen methods allow for the collection and analysis of data in a manner that respects participant confidentiality and research integrity, in line with guidelines from the University of Turku and the Finnish National Board on Research Integrity (TENK, 2019). The data management plan is added as appendix E.

All participant responses, including survey data, detection performance scores, and psychological test results, are stored anonymously. Data is collected via Qualtrics, which complies with GDPR standards and provides secure, encrypted transmission and storage (Qualtrics, 2021).

No identifying information such as names, email addresses, or IP addresses will be stored or used in analysis. From an ethical standpoint, psychological discomfort may arise from completing personality or stress-related assessments. This is mitigated by ensuring the voluntary nature of participation, offering the option to withdraw at any time, and avoiding sensitive or intrusive questions (Price, 2015). All participation is confidential and anonymous, and no results are shared with employers or third parties. The participants are informed by the participants briefing illustrated in appendix D.

4 Data analysis and findings

In this study, a dual analytical approach with both multiple regression analysis and fsQCA is employed to comprehensively investigate how cognitive and psychological factors influence employees' deepfake detection accuracy under time pressure. Multiple regression analysis, a symmetric method, estimates the direct (partial) effects of each predictor, providing insights into the unique contribution of each variable to detection performance. However, it assumes that relationships between variables are linear, additive, and consistent across all cases. In contrast, human cognition, psychological traits, and organizational behavior are often complex, non-linear, and context-dependent (Cohen et al., 2003).

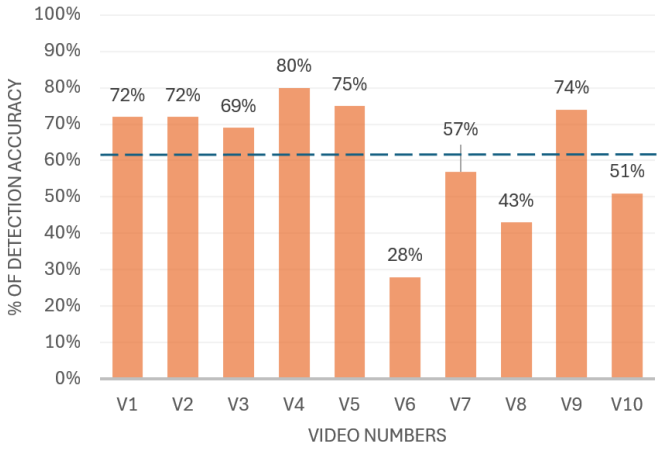
Asymmetric methods like fsQCA do not rely on assumptions of linearity or additivity. Instead, they aim to identify combinations of causal conditions that can lead to the same outcome, acknowledging that multiple, distinct pathways may result in successful deepfake detection (Pappas & Woodside, 2021; Ragin, 2000). Unlike crisp-set QCA (csQCA) that requires binary conditions (e.g., "high/low"), fsQCA accommodates continuous or graded membership in sets (Ragin, 2009). This aligns with psychological constructs measured on a likert scale and continuous variables. By combining these two methodological approaches (Kaya et al., 2020; Sheng et al., 2025), this study seeks to provide a richer, more nuanced understanding of how cognitive, psychological, and experience variables interact to shape employees' ability to detect deepfakes under time pressure.

4.1 Descriptive statistics

The respondents ability to detect deepfakes varied widely across different video stimuli, with accuracy rates ranging from 28% to 80%, indicating that some videos were considerably more difficult to judge than others (see Figure 4). The overall detection accuracy was 62% (SD = 15%), suggesting a moderate ability to distinguish real from fake content. This heterogeneity in detection accuracy rates highlights that the experimental design successfully captured a realistic spectrum of deepfake sophistication, thereby enhancing the ecological validity of the study (Groh et al., 2021).

Figure 4

Deepfake detection accuracy by video sample number

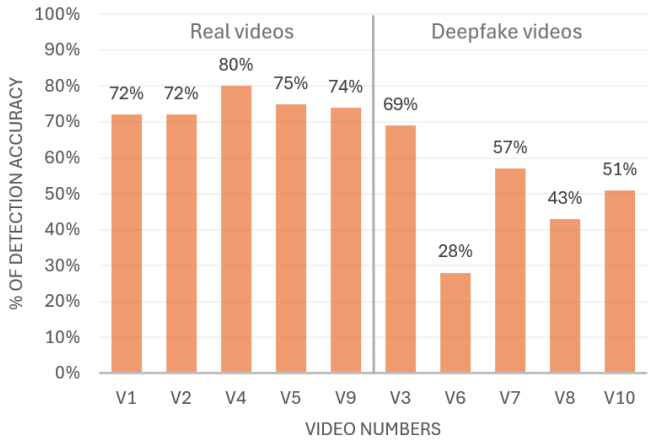


Note. This bar chart displays the percentage of correct deepfake detections per video (V1–V10). Detection accuracy ranges from a high of 80% (V4) to a low of 28% (V6). The dashed horizontal line represents the overall average detection accuracy across all video (62%).

On average, participants correctly identified authentic (real) videos approximately 75% of the time, while their accuracy in detecting deepfakes was substantially lower, at around 48%, as illustrated in Figure 5. This discrepancy reflects a well-documented "truth bias" or "detection bias," wherein individuals tend to assume content is genuine unless there are explicit cues indicating deception (Köbis et al., 2021; Levine, 2022). This bias is grounded in Truth-Default Theory, which posits that people are predisposed to accept messages as truthful in the absence of compelling reasons for doubt (Levine, 2014). Consequently, individuals are more proficient at recognizing authentic footage than at identifying sophisticated manipulations such as deepfakes.

Figure 5

Comparison of detection accuracy for real versus deepfake videos across video samples



Note. This bar chart compares participants' detection accuracy between real and deepfake video samples. The real videos (V1, V2, V4, V5, V9) consistently show high detection accuracy, ranging from 72% to 80%. In contrast, detection performance on deepfake videos (V3, V6, V7, V8, V10) varies widely, with accuracy dropping as low as 28% (V6).

Table 2. Descriptive statistics

Variable	Mean	SD	Min	Max
Deepfake Detection Accuracy	0.62	0.15	0.30	0.90
Cognitive workload (NASA-TLX)	9.03	3.34	1.33	16.00
Sense of coherence (SOC-13)	4.78	0.66	3.32	6.18
Extraversion	3.34	0.88	1.00	5.00
Agreeableness	3.65	0.76	1.50	5.00
Conscientiousness	3.76	0.86	1.00	5.00
Neuroticism	2.67	0.99	1.00	5.00
Openness	3.43	0.92	1.00	5.00
Detection Experience	2.67	1.25	1.00	5.00

Table 2 summarizes key descriptive statistics for all study variables. On average, participants correctly identified 62 % of deepfakes (SD = 0.15; range = 0.30–0.90), and reported a moderate cognitive workload on the NASA-TLX (M = 9.03, SD = 3.34; range = 1.33–16.00). Sense of coherence scores (SOC-13) clustered around the upper midpoint (M = 4.78, SD = 0.66; range =

3.32–6.18), indicating generally strong perceptions of manageability and meaning. Personality traits fell near their scale midpoints: extraversion (M = 3.34, SD = 0.88), agreeableness (M = 3.65, SD = 0.76), conscientiousness (M = 3.76, SD = 0.86), neuroticism (M = 2.67, SD = 0.99), and openness (M = 3.43, SD = 0.92). Finally, prior deepfake detection experience averaged 2.67 on a 1–7 scale with a maximum score of 5 (SD = 1.25; observed range = 1.00–5.00), showing that participants had moderate detection experience at maximum.

4.2 Multiple regression analysis

To explain how various cognitive, psychological, and experiential factors influence employees' ability to detect deepfakes under time pressure, a multiple linear regression analysis is conducted. This approach allows for the simultaneous examination of several predictors to assess their individual contributions to detection accuracy while controlling for the effects of others (Field, 2018). The regression model includes measures of personality traits, sense of coherence, cognitive workload, and prior detection experience. Given the explanatory nature of this study and the theoretical complexity of human behavior, this analysis serves as a foundational step in identifying which individual factors are statistically significant and practically meaningful predictors of deepfake detection performance.

4.2.1 Data preprocessing

The collected survey responses and deepfake video classification results together constitute the study's dataset. Before analysis, this dataset undergoes a structured preprocessing phase to ensure data quality and consistency (AlSobeh, 2024). This process includes data cleaning, such as removing incomplete or invalid responses, checking for out-of-range values, and addressing missing data, followed by the extraction of relevant features from both questionnaire responses and detection task outcomes (Mirzaei et al., 2022).

4.2.2 Post hoc G*power analysis

A post hoc power analysis for the multiple regression predicting deepfake detection accuracy (R^2 = .216) was conducted in G*Power (Version 3.1; Faul et al., 2009). The effect size f^2 was computed as f^2 =0.2755, which falls in the medium-to-large range (Cohen, 1988; Erickson, 2017). The following settings were used: F tests; linear multiple regression (fixed model, R^2 deviation from zero); post hoc analysis; $f^2 = 0.2755$; $\alpha = .05$; N = 89; and eight predictors. G*Power reported a noncentrality parameter of $\lambda = 24.52$, a critical $F(8, 80) = 2.056$, and achieved power $(1 – \beta) = .945$.

These results exceed the conventional 0.80 threshold for adequate power (Cohen, 1988; Faul et al., 2009; Raywood-Burke, 2023), indicating a low probability of a Type II error and confirming that the sample size was sufficient to detect the observed effect. The critical F value of 2.056 marks the rejection boundary for the null hypothesis at $\alpha = .05$, and the high noncentrality parameter underscores the sensitivity of the test given the specified effect size and sample size (Cohen, 1988). Overall, the analysis reinforces confidence in the regression findings.

4.2.3 Measurement model

The internal consistency of the multi-item scales was assessed using Cronbach's alpha. Most scales showed acceptable to good reliability, particularly the SOC-13 scale ($\alpha = .828$). However, some personality subscales (e.g., agreeableness, $\alpha = .313$) showed low internal consistency, which affects the reliability of their effects in the regression models. The Cronbach alpha score per variable is illustrated in Table 3. Research on brief personality inventories, like the BFI-10, highlights trade-offs between efficiency and psychometric quality. Short scales often show lower internal consistency and inter-item correlations compared to their longer counterparts (Park et al, 2022).

Table 3. Cronbach's alpha value per variable

Variable	Cronbach alpha
Cognitive workload (NASA-TLX)	.580
Sense of coherence (SOC-13)	.828
Extraversion	.510
Agreeableness	.313
Conscientiousness	.510
Neuroticism	.646
Openness	.681

4.2.4 Structural model

The model's coefficient of determination was $R^2 = .216$ (adjusted $R^2 = .137$), meaning the predictors explain about 21.6% of the variance in detection accuracy. R^2 quantifies the goodness-of-fit of the regression, so an R^2 of .216 reflects only a modest fit: roughly one-fifth of the outcome

variability is explained by the model. In practical terms, this implies that 78% of the variance remains unexplained, suggesting other factors influence detection accuracy (Field, 2018). However, in social-science research, relatively low R^2 values (on the order of 0.10–0.20) are often considered acceptable (Ozili, 2023). Thus, while the model's predictors have a statistically reliable effect, the modest R^2 indicates that the model captures only a limited portion of the total variance in deepfake detection performance.

Multicollinearity was evaluated using the Variance Inflation Factor (VIF). Following Hair et al. (2010), VIF values below five are considered acceptable, indicating that predictors do not exhibit problematic levels of shared variance. As shown in Table 4, all variables met this criterion, with VIF values ranging from 1.05 to 1.68, suggesting no serious multicollinearity issues.

Table 4. VIF per variable

Variable	VIF
Cognitive workload (NASA-TLX)	1.05
Sense of coherence (SOC-13)	1.68
Extraversion	1.37
Agreeableness	1.25
Conscientiousness	1.32
Neuroticism	1.59
Openness	1.13
Detection Experience	1.27

4.2.5 Effect analysis

Of the eight predictors examined, only detection experience demonstrated a robust positive association with deepfake detection accuracy ($b = 0.0425$, $SE = 0.012$, $t = 3.45$, $p = .001$, 95% CI [0.018, 0.067]), indicating that greater familiarity with deepfakes significantly enhances one's ability to recognize them under time pressure.

All other variables yielded non-significant effects: cognitive workload (NASA-TLX) exerted an almost null influence ($b = 0.0017$, $p = .700$, 95% CI [–0.007, 0.011]), sense of coherence showed a slight positive but non-significant trend ($b = 0.0386$, $p = .180$, 95% CI [–0.018, 0.095]), and the Big

Five traits, extraversion (b = –0.0067, p = .731, 95% CI [–0.046, 0.032]), agreeableness (b = –0.0089, p = .681, 95% CI [–0.052, 0.034]), conscientiousness (b = 0.0185, p = .343, 95% CI [–0.020, 0.057]), neuroticism (b = 0.0274, p = .139, 95% CI [–0.009, 0.064]), and openness (b = 0.0238, p = .160, 95% CI [–0.010, 0.057]), did not meet conventional significance thresholds. These results suggest that, within this sample, prior detection experience is the primary driver of deepfake identification performance.

Table. 5 Predictors of deepfake detection accuracy (N = 89)

Variable	Coefficient	Standard Error	t-value	p-value	95% Confidence Interval
Constant	0.0998	(0.182)	0.548	0.585	[-0.262, 0.462]
Cognitive workload (NASA-TLX)	0.0017	0.004	0.387	0.700	[–0.007, 0.011]
Sense of coherence (SOC13)	0.0386	0.028	1.354	0.180	[–0.018, 0.095]
Extraversion	–0.0067	0.020	–0.345	0.731	[–0.046, 0.032]
Agreeableness	–0.0089	0.022	–0.412	0.681	[–0.052, 0.034]
Conscientiousness	0.0185	0.019	0.953	0.343	[–0.020, 0.057]
Neuroticism	0.0274	0.018	1.496	0.139	[–0.009, 0.064]
Openness	0.0238	0.017	1.418	0.160	[–0.010, 0.057]
Detection Experience***	0.0425	0.012	3.454	0.001	[0.018, 0.067]

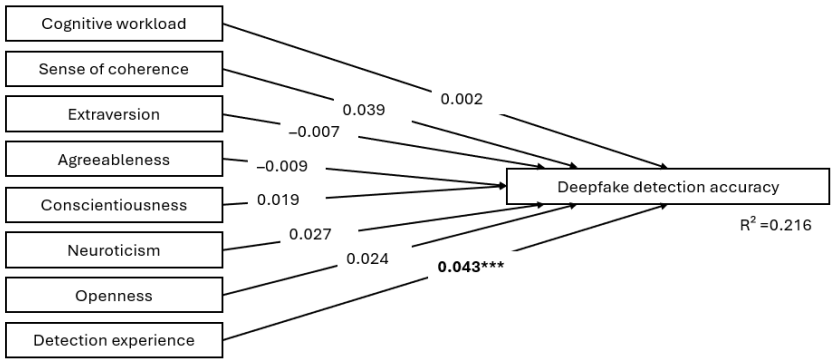
Note.* Standard errors in parentheses. *, *, * indicate significance at, 90%, 95%, and 99% levels respectively.

***Note.* Regression model predicting deepfake detection accuracy with N = 89; $F(8, 80) = 2.752$, $p = 0.00975$; $R^2 = 0.216$; Adj. $R^2 = 0.137$

Figure 6 visualizes the results of a multiple regression analysis predicting deepfake detection accuracy based on psychological, cognitive, and experience variables. Each arrow represents a predictor, with the accompanying coefficient indicating the strength and direction of its relationship. This visual representation supports the interpretation that practical experience outweighs personality or cognitive workload in explaining deepfake detection performance.

Figure 6.

Predictors of deepfake detection accuracy (regression coefficients)



Note. This conceptual model illustrates the relationship between various human vulnerability factors and deepfake detection accuracy. Standard errors in parentheses. *, *, *** indicate significance at 90%, 95%, and 99% levels respectively. Regression model predicting deepfake detection accuracy with N = 89; F(8, 80) = 2.752, p = 0.00975; R^2 = 0.216; Adj. R^2 = 0.137

4.3 Fuzzy-set qualitative comparative analysis

To address the limitations of symmetric regression analysis and capture the complex interplay of cognitive and psychological factors, this study employs fsQCA. This asymmetric approach identifies how combinations of conditions, rather than isolated variables (Ragin, 2009), collectively influence deepfake detection accuracy under time pressure.

Several steps are required to conduct the fsQCA analysis. First, the reliability and validity of the variables are assessed, which has already been done within the multiple linear regression analysis. Second, calibration thresholds are determined based on the guidelines provided by Pappas and Woodside (2021). For variables measured on a 5-point Likert scale (e.g., personality traits), the thresholds are set at 2 (full non-membership), 3 (crossover point), and 4 (full membership). For 7-point Likert scale items (e.g., SOC-13), the thresholds are set at 2 (non-membership), 4 (crossover), and 6 (full membership). For all other continuous variables, the thresholds are calibrated at the 0.05 percentile (full non-membership), the 0.50 percentile (crossover point), and the 0.95 percentile (full membership), as illustrated in Table 5. Following the recommendation of Fiss (2011), a constant of 0.001 is added to accommodate intermediate-set/crossover calibration.

Third, a truth table is constructed, which is a table that lists all possible combinations of conditions. Each row represents a unique configuration of conditions. The truth table gets sorted by frequency and consistency. The fourth step is to obtain the solution; in this context, a "solution" denotes a group of condition combinations that are consistently associated with the outcome and are backed by a substantial number of cases demonstrating this relationship. An intermediate solution is chosen as this is the default choice for causal analysis (Pappas & Woodside, 2021). The findings are then illustrated in a table.

Interpretation of fsQCA results proceeds in three steps. First, weak solutions are filtered out, leaving only those with sufficient coverage to demonstrate empirical relevance. Second, sufficiency analysis uncovers “recipes” (configurations of present and absent conditions) that consistently produce the outcome. Each configuration’s consistency and coverage then indicate its reliability and explanatory reach. Finally, by comparing raw and unique coverage, it becomes clear which pathways offer their distinct explanations, helping to capture the full complexity of how causes combine (Pappas & Woodside, 2021).

Both high deepfake detection accuracy and low deepfake detection accuracy are treated as separate fuzzy sets to honor causal asymmetry: factors driving success need not simply invert those driving failure (Pappas & Woodside, 2021). By preserving nuanced degrees of set membership, this dual-set approach uncovers unique configurational pathways to both high and low detection accuracy outcomes, offering deeper insights than conventional symmetric net-effects models (Olya & Altinay, 2016).

4.3.1 Calibrating variables

Following established fsQCA guidelines, the calibration of fuzzy sets should rely on both substantive knowledge and empirical data, rather than merely applying scale endpoints in a mechanical or theoretical way. As Ragin (2009) points out, the process of calibration must align with how the researcher conceptually defines and labels the set under study.

In the context of this study, the calibration of deepfake experience was adjusted accordingly: since no participants reported having extensive or very extensive experience, the maximum observed level was "moderate." Therefore, the upper calibration threshold was set at this realistic maximum, reflecting the actual distribution of experience within the sample. All calibrations are illustrated in Table 6.

Table 6. fsQCA calibration thresholds table

Variable	Full Non-Membership (0.05)	Crossover Point (0.50)	Full Membership (0.95)
deepfake_detection_accuracy	0.40	0.60	0.86
Cognitive workload (NASA-TLX)	3.19	9.07	13.55
Sense of coherence (SOC-13)	2	4	6
Extraversion	2	3	4
Agreeableness	2	3	4
Conscientiousness	2	3	4
Neuroticism	2	3	4
Openness	2	3	4
Detection Experience	2	3	4

4.3.2 FsQCA solutions for high detection accuracy

The fsQCA analysis identified four distinct configurations ("recipes") leading to high deepfake detection accuracy. The intermediate solution incorporates a subset of the simplifying assumptions used in the parsimonious solution, guided by theoretical and empirical knowledge (Fiss, 2011).

The configurations are characterized by conditions deemed either necessary or sufficient, with distinctions made between core and peripheral elements. Core conditions refer to fundamental components that consistently appear across configurations and demonstrate a strong causal link to the outcome. Peripheral conditions have a weaker causal relation to the outcome. By comparing the intermediate and parsimonious solutions, core conditions can be distinguished, as the parsimonious solution excludes all peripheral elements (Pappas & Woodside, 2021).

In Table 7, a filled circle (●) indicates the presence of a condition at a high level, while a crossed-out circle (Ø) indicates its explicit absence (low level). A blank cell represents a "do not care" condition, meaning it can be either high or low without affecting the configuration's outcome (Pappas & Woodside, 2021). The extended truth table analysis is added as Appendix F. For example, Configuration 3 includes low extraversion, low agreeableness, high openness, and a high sense of coherence, with the following core conditions: high conscientiousness, high neuroticism, and high cognitive workload.

Table 7. Investigation of high deepfake detection accuracy

	Solutions			
Configuration	1	2	3	4
Extraversion	○	●	○	●
Agreeableness	●	●	○	●
Conscientiousness	●	Ø	⬤	●
Neuroticism	○	⬤	⬤	⬤
Openness	●	●	●	●
Cognitive workload	○	Ø	⬤	○
Sense of coherence	●	●	●	●
Detection experience		Ø		●
Raw Consistency	.887	.928	.958	.977
PRI Consistency	.757	.711	.797	.942
Raw coverage	.180	.108	.119	.172
Unique coverage	.045	.015	.039	.069
Overall solution consistency		.797		
Overall solution coverage		.570		

**Note 1.* A black circle (●) signifies that a condition is present, while a white circle with a mark (○) indicates that a condition is absent. Large circle Ø / ⬤, core condition; Small circle ○ / ● peripheral condition; Blank space 'do not care' condition.

***Note 2.* Filters PRI consistency > 0.7 and Raw consistency > 0.85

After applying the recommended raw consistency threshold of 0.85 (Greckhamer et al., 2013), all four solution configurations demonstrated very high consistency, with raw consistency values ranging from 0.887 to 0.977 and PRI consistency values from 0.711 to 0.942. This indicates that each configuration is strongly associated with high deepfake detection accuracy (Ragin, 2009).

Configuration 1 is characterized by low extraversion, low neuroticism, and low cognitive workload, combined with high agreeableness, openness, and a high sense of coherence, together leading to high deepfake detection accuracy. In Configuration 2, a unique pattern emerges, featuring three absent core conditions: conscientiousness, cognitive workload, and detection experience. These are combined with high extraversion, agreeableness, openness, a strong sense of coherence, and high neuroticism as a core condition. Configuration 3 includes low extraversion, low agreeableness, high openness, and a high sense of coherence, with the following core conditions: high conscientiousness, high neuroticism, and high cognitive workload.

Configuration 4, characterized by high extraversion, high agreeableness, high conscientiousness, and high neuroticism as core conditions, along with high openness, a high sense of coherence, high detection experience, and low cognitive workload, stood out as the most robust example. With a raw consistency of 0.977, nearly all cases exhibiting this combination show high accuracy. Its PRI consistency of 0.942 further confirms that the configuration is reliably sufficient and largely free of contradictory cases. The fact that all PRI values exceed 0.7 reinforces the reliability of these configurations in explaining the outcome (Greckhamer et al., 2018).

By comparing the various solutions, several notable patterns emerge. For instance, neuroticism appears as a core present condition in solutions 2-4, highlighting its central role in configurations associated with higher deepfake detection accuracy. Although classified as peripheral, the personality trait openness is consistently present across all solutions, suggesting a meaningful, but less decisive, contribution. Sense of coherence follows a similar pattern, reinforcing its potential relevance despite not being a core condition. However, many variables do not exhibit a consistent presence or absence across the configurations. This also demonstrates the principle of equifinality, that multiple pathways can lead to the same outcome (Schneider & Wagemann, 2012).

Raw coverage values for each configuration (ranging from 0.108 to 0.180) indicate the proportion of high-accuracy cases explained by each pathway (Pappas & Woodside, 2021). Configuration 1 shows the highest raw coverage (0.180), accounting for 18% of the high-accuracy cases. However, unique coverage values are low (ranging from 0.015 to 0.069), suggesting that no single configuration accounts for a large proportion of cases on its own (Pappas & Woodside, 2021).

4.3.3 FsQCA solutions for low detection accuracy

Analyzing the low-Y solutions reveals patterns of absent traits and conditions that contribute to poor deepfake detection accuracy. Given the asymmetric nature of fsQCA, the pathways leading to

high accuracy cannot simply be inverted to explain failure (Ragin, 2009). Including the low-Y configurations thus offers a more comprehensive causal picture by identifying which solutions are sufficient for low detection performance, thereby enhancing both theoretical insights and practical implications (Olya & Altinay, 2016). The extended truth table analysis is added as Appendix G.

Table 8. Investigation of low deepfake detection accuracy

	Solutions	
Configuration	1	2
Extraversion	○	●
Agreeableness	○	●
Conscientiousness	●	○
Neuroticism	○	○
Openness	Ø	Ø
Cognitive workload	Ø	Ø
Sense of coherence	●	●
Detection experience	Ø	Ø
Raw Consistency	.974	.991
PRI Consistency	.915	.957
Raw coverage	.115	.107
Unique coverage	.021	.015
Overall solution consistency		.791
Overall solution coverage		.506

**Note 1.* A black circle (●) signifies that a condition is present, while a white circle with a mark (○) indicates that a condition is absent. Large circle Ø / ●, core condition; Small circle ○ / ● peripheral condition; Blank space 'do not care' condition.

***Note 2.* Filters PRI consistency > 0.7 and Raw consistency > 0.85

The consistency thresholds were aligned with those established for the high-accuracy solutions, following the criteria recommended by Greckhamer et al. (2018). Specifically, a minimum raw

consistency of 0.85 and a PRI consistency of at least 0.70 were applied. Two configurations met both thresholds, with raw consistency values ranging from 0.974 to 0.991 and PRI consistency values between 0.915 and 0.957. These results indicate that the identified configurations, primarily characterized by the absence of conditions, consistently explain instances of low deepfake detection accuracy.

The raw coverage values, which indicate the proportion of low-accuracy cases explained by each configuration (Ragin, 2009), were 0.115 and 0.107, respectively. However, the unique coverage values were relatively low (0.021 and 0.015), suggesting that each configuration explains a portion of the outcome that partly overlaps with the other.

Solutions 1 and 2 demonstrate high consistency. Solution 1 shows a raw consistency of 0.974 and PRI consistency of 0.915, while Solution 2 has a raw consistency of 0.991 and PRI consistency of 0.957. The pathways have notable similarities; both are characterized by the absence of neuroticism and the absence of three core conditions: openness, cognitive workload, and detection experience. However, they differ in the presence or absence of extraversion, agreeableness, and conscientiousness. Sense of coherence is present in both configurations. These pathways result in low deepfake detection accuracy.

5 Discussion

This study investigated the cognitive and psychological factors that influence employees' deepfake detection accuracy under time pressure, addressing a critical gap in cybersecurity research at the intersection of human factors and synthetic media threats. The findings reveal both expected and surprising patterns that contribute to our understanding of human vulnerabilities in deepfake-based social engineering attacks. Both linear relationships and configurational pathways were tested using multiple regression analysis and fsQCA.

Multiple regression analysis and fsQCA were employed to examine the proposed model. In support of H1, the multiple regression results indicated that prior deepfake detection experience significantly enhances detection accuracy (b = 0.0425, p = .001), aligning with the Elaboration Likelihood Model's concept of peripheral route processing under time pressure (Goh, 2023). Detection experience appears to strengthen the ability to apply media-based identification strategies, such as recognizing graphical anomalies and production quality flaws commonly associated with synthetic media (Goh et al., 2023). This finding underscores the importance of training and familiarity in developing detection competencies under time-constrained conditions (Somoray & Miller, 2023; Diel et al., 2024).

Contrary to expectations outlined in H2a–H2e, none of the Big Five personality traits demonstrated significant linear effects on deepfake detection accuracy. This outcome contrasts with earlier research in social engineering and phishing contexts, where traits such as extraversion and agreeableness have been consistently linked to increased susceptibility (Ei Bolock & Madany, 2025) or conscientiousness with decreased susceptibility (Uebelacker et al., 2014). Concerning H3, sense of coherence exhibited a slight positive but statistically non-significant effect, suggesting only a limited influence on detection outcomes in the tested model. Finally, H4 was also not supported, as cognitive workload showed a negligible and non-significant association with deepfake detection accuracy. This result contradicts theoretical assumptions that higher cognitive workload would impair performance under time pressure (Dsouza et al., 2024; Montanez et al., 2020).

The fsQCA revealed the complexity underlying human deepfake detection performance. Four distinct configurations emerged as sufficient for high detection accuracy, demonstrating the principle of equifinality, that multiple pathways can lead to the same outcome (Schneider & Wagemann, 2012). In line with Proposition 1, these configurations illustrate that different combinations of detection experience, personality traits, sense of coherence, and cognitive workload

can result in effective detection performance. Among these, Configuration 4 emerged as the most robust, characterized by high levels of extraversion, agreeableness, conscientiousness, and neuroticism as a core condition, in combination with high detection experience and low cognitive workload (raw consistency = 0.977, PRI consistency = 0.942). The presence of neuroticism as a core condition in solutions 2 through 4 was unexpected, as its directionality appears to contradict prior theoretical assumptions (Halevi et al., 2016). Openness and sense of coherence were present in all higher deepfake detection scores as peripheral conditions.

Supporting Proposition 2, two configurations were associated with low detection accuracy. Both were defined by the absence of neuroticism as a peripheral condition and the absence of openness, cognitive workload, and detection experience as core conditions. The absence of detection experience as a core condition in lower detection scores corresponds with prior theoretical assumptions and empirical findings. Whilst other personality traits like extraversion and conscientiousness were peripheral and inconsistent, as shown in Table 8. The commonalities within the two solutions (e.g., absence of openness) could give guidance for future research.

These findings underscore the configurational nature of deepfake detection, where different pathways, rather than individual predictors, shape successful or unsuccessful detection outcomes. This reflects the core principles of configurational analysis: conjunctural causation (conditions matter only in combination), equifinality (multiple distinct configurations can lead to the same outcome), and causal asymmetry (the pathways to success are not simply the reverse of those leading to failure) (Ragin, 2009).

5.1 Theoretical implications

The study's findings provide significant theoretical extensions to the ELM in the context of deepfake detection. The results demonstrate that under time pressure conditions, individuals rely on peripheral processing routes, leading to media-based identification strategies for detecting deepfakes (Goh, 2023). The strong empirical support for detection experience as the primary predictor of accuracy (b = 0.0425, p = .001) aligns with ELM's prediction that peripheral cues become more influential when cognitive resources are limited. This can be interpreted as an experience-based heuristic acting as a peripheral cue. Individuals with hands-on familiarity recognize subtle production cues or flaws. In other words, time-pressured participants defaulted to a “seeing-is-believing” heuristic (mistaking deepfakes for real) unless prior experience alerted them to deception.

Second, this study enriches social-engineering and personality theories by revealing the configurational complexity of human vulnerability. Traditional social-engineering models often examine individual factors (e.g., personality traits) in isolation, but findings have been inconsistent (Abraham et al., 2023; Halevi et al., 2016; Lawson et al., 2018). Montañez et al. (2020) and Wang et al. (2021) noted that personality traits influence susceptibility, yet no single trait pattern emerges. Standard regression showed none of the personality traits had a direct effect, echoing prior deepfake detection studies that also found weak or null links between personality and detection accuracy (Abraham et al., 2023). Instead, fsQCA uncovered multiple distinct pathways (configurations of personality traits, workload, sense of coherence, and detection experience) that lead to high/low accuracy. For example, high neuroticism, typically viewed as a vulnerability factor in phishing, emerged as a protective factor. In this context, anxious individuals may be more skeptical and thus more inclined to assume that videos are deepfakes rather than real. This aligns with the concept of truth bias, as anxiety-related traits can heighten suspicion and reduce default trust (Levine, 2022). Likewise, openness to experience (intellectual curiosity and flexibility) appeared as a core absent condition in low detection recipes. This implies that individuals high in openness, due to their willingness to consider unconventional cues or possibilities, may be more inclined to question apparent authenticity. In effect, the results provide evidence that no ‘one-size-fits-all’ trait profile predicts susceptibility, but multiple distinct profiles can yield high and low detection accuracy. This has implications for theory and practice. Models of human vulnerability must account for equifinality and causal asymmetry, echoing calls by Ragin (2009) and Fiss (2011) to move beyond linear assumptions.

Third, this research contributes to cognitive workload and stress-related theory in information processing. Conventional cognitive workload theory predicts that high workload degrades performance (Plass et al., 2010), and social-engineering frameworks suggest that stress and cognitive overload increase susceptibility. The linear analysis found no main effect of cognitive workload or sense of coherence on detection accuracy. This suggests that under time pressure, people may adopt compensatory strategies or shift to peripheral processing in ways that basic workload models do not anticipate. In other words, stress resilience (sense of coherence) or cognitive workload alone did not determine performance. Instead, the essence of both variables is notable only via combinations with other factors. This opens a theoretical gap. Existing models of cognitive load and stress in cybersecurity may need refinement to account for interactive effects. For instance, Montañez et al. (2020) highlighted high cognitive workload as a key vulnerability factor. However, the results of this thesis suggest that low cognitive workload is associated with

poor detection performance, particularly when combined with low openness to experience or a lack of prior detection experience. By exposing these nuanced patterns, this thesis extends cognitive workload theory.

From a methodological standpoint, this study makes an important contribution by combining two different but complementary research approaches. As recommended by Pappas & Woodside (2021), the study blends traditional regression analysis with fsQCA. This combination uncovered both straightforward relationships and complex patterns that influence how people detect deepfakes under time pressure. To the best of the author's knowledge, no existing social engineering studies on human vulnerability employ this methodological combination, making its use in this study a novel contribution. Future research could benefit from this integrative approach to better capture both linear effects and configurational complexity.

5.2 Managerial implications

The empirical results demonstrate that detection experience is the single strongest and most reliable predictor of employees' ability to detect deepfakes under time pressure. Operationally, this finding implies that deepfake detection should be treated by managers as a trainable and strategically critical competency (Diel et al., 2024).

To strengthen training initiatives within organizations, Article 4 of the EU AI Act can serve as a regulatory lever (Regulation (EU) 2024/1689, art. 4.). Managers can use this policy framework to, for example, secure dedicated budgets for detection training, define training KPIs (e.g., percentage of employees achieving ≥ 75% detection accuracy during quarterly drills) as auditable compliance evidence, and mandate onboarding literacy modules for all new hires. In doing so, they effectively align HR practices with compliance obligations (Cetindamar et al., 2022).

Several existing frameworks can support this effort, notably the DigComp 2.2 framework developed by the European Commission. This framework provides a structured approach for building AI- and information-literacy skills among people, enabling organizations to design compliant and futureproof training programs (Vuorikari et al., 2022).

Organizations can enhance the effectiveness of security awareness training by tailoring content and delivery methods to employees' levels of detection experience and personality traits, particularly openness and neuroticism. Following the framework proposed by Alotaibi et al. (2023), this approach enables the development of personalized security awareness programs that are more responsive to individual differences.

5.3 Limitations and future research

Several limitations should be acknowledged when interpreting these findings. The modest explanatory power of the regression model ($R^2 = .216$) indicates that approximately 78% of the variance in detection accuracy remains unexplained. Lower R^2 values are considered acceptable (Ozili, 2023); still, this suggests that important factors influencing deepfake detection were not captured in the current study. Future research should explore additional variables that may contribute to detection performance.

The personality trait agreeableness yielded a Cronbach's alpha of .313, which is considered unacceptably low, indicating poor internal consistency of the scale. This limitation may stem from the brevity of the BFI-10. A more comprehensive instrument, such as the BFI-44, could address these internal consistency issues by offering a richer and more reliable measurement of the trait (Rammstedt & John, 2006).

The reliance on subjective measures for cognitive workload (NASA-TLX) rather than objective physiological indicators may have limited the sensitivity of this variable (Kosch et al., 2023). Future studies should consider incorporating objective measures such as electroencephalogram (EEG) or transcranial Doppler blood flow measurements to provide more precise assessments of cognitive workload (Khan et al., 2023).

The sample characteristics also present limitations. With 89 participants recruited primarily through personal networks, the generalizability of findings may be limited (Tipton et al., 2017; Goodman, 2011). Future research should employ larger, more diverse samples that include individuals from various organizational contexts and cultural backgrounds. Lastly, longitudinal research designs would provide valuable insights into how detection skills develop over time and whether the effects of experience remain stable across different threat landscapes.

6 Conclusions

This study investigated the psychological and cognitive factors that influence employees' ability to detect deepfakes under time pressure, addressing a critical gap at the intersection of human vulnerabilities and cybersecurity. By combining multiple regression and fsQCA, the research provided both symmetric and configurational insights into the factors contributing to successful deepfake detection.

Key findings from the multiple regression analysis revealed that prior deepfake detection experience was the only significant predictor of detection accuracy under time constraints; none of the measured personality traits, sense of coherence, or cognitive workload showed direct, statistically significant effects on detection performance. In contrast, the fsQCA uncovered multiple distinct configurations that led to high or low detection accuracy, underscoring the equifinality of deepfake detection success or failure. With the presence of core condition, neuroticism leading to high accuracy (solutions 2-4), and the absence of core conditions, openness, cognitive workload, and detection experience consistently characterized low detection accuracy.

Practically, organizations should prioritize hands-on deepfake training. Future research should aim to increase sample diversity, use more comprehensive personality and resilience measures, and explore the long-term impact of training to enhance generalizability and applicability. Ultimately, integrating tailored human-centred strategies with technological solutions will be essential to fortify organizations against the evolving threat of deepfake-based social engineering.

References

Aaron, L. J. (2020). Dollars, deception, and deepfakes: An analysis of deepfakes and synthetic media fraud [Doctoral dissertation, Utica College]. ProQuest Dissertations Publishing. https://www.proquest.com/dissertations-theses/dollars-deception-deepfakes-analysis-synthetic/docview/2414420537/se-2

Abraham, J., Putra, H. A., Prayoga, T., Warnars, H. L. H. S., Manurung, R. H., & Nainggolan, T. (2023). Prediction of self-efficacy in recognizing deepfakes based on personality traits. *F1000Research*, *11*, 1529.

Ahmed, S. (2021). Fooled by the fakes: Cognitive differences in perceived claim accuracy and sharing intention of non-political deepfakes. Personality and Individual Differences, 182, 111074. https://doi.org/10.1016/j.paid.2021.111074

Alotaibi, S., Furnell, S., & He, Y. (2023, July). Towards a framework for the personalization of cybersecurity awareness. In *International Symposium on Human Aspects of Information Security and Assurance* (pp. 143-153). Cham: Springer Nature Switzerland.

AlSobeh, A., Franklin, A., Woodward, B., Porche, M., & Siegelman, J. (2024). Unmasking media illusion: analytical survey of deepfake video detection and emotional insights. *Issues in Information Systems*, *25*(2).

Altuncu, E., Franqueira, V. N. L., & Li, S. (2024). Deepfake: definitions, performance metrics and standards, datasets, and a meta-review. *Frontiers in Big Data*, *7*. https://doi.org/10.3389/fdata.2024.1400024

Antonovsky, A. (1993). The structure and properties of the sense of coherence scale. *Social science & medicine*, *36*(6), 725-733

Ask, T. F., Lugo, R., Fritsch, J., Veng, K., Eck, J., Özmen, M., Bärreiter, B., Knox, B. J., & Sütterlin, S. (2023). *Cognitive flexibility but not cognitive styles influence deepfake detection skills and metacognitive accuracy*. https://doi.org/10.31234/osf.io/a9dwe

Bateman, J. (2022). *Deepfakes and synthetic media in the financial system: Assessing threat scenarios*. Carnegie Endowment for International Peace.

Bodepudi, A., & Reddy, M. (2020). Spoofing attacks and mitigation strategies in biometrics-as-a-service systems. Eigenpub Review of Science and Technology, 4(1), 1-14.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101.

Bray, S. D., Johnson, S. D., & Kleinberg, B. (2023). Testing human ability to detect 'deepfake' images of human faces. *Journal of Cybersecurity*, *9*(1). https://doi.org/10.1093/cybsec/tyad011

Bryman, A. (2016). Social Research Methods (5th ed.). Oxford University Press.

Butavicius, M., Taib, R., & Han, S. J. (2022). Why people keep falling for phishing scams: The effects of time pressure and deception cues on the detection of phishing emails. *Computers & Security*, *123*, 102937. https://doi.org/10.1016/j.cose.2022.102937

Byrne, K. A., Silasi-Mansat, C. D., & Worthy, D. A. (2014). Who chokes under pressure? The Big Five personality traits and decision-making under pressure. *Personality and Individual Differences*, *74*, 22–28. https://doi.org/10.1016/j.paid.2014.10.009

Cacioppo, R. E. Petty, C. F. Kao, and R. Rodriguez, ''Central and peripheral routes to persuasion: An individual difference perspective,'' J. Personality Social Psychol., vol. 51, no. 5, p. 1032, 1986.

Cascio, W. F., & Montealegre, R. (2016). How technology is changing work and organizations. *Annual Review of Organizational Psychology and Organizational Behavior*, *3*(1), 349–375. https://doi.org/10.1146/annurev-orgpsych-041015-062352

Cetindamar, D., Kitto, K., Wu, M., Zhang, Y., Abedin, B., & Knight, S. (2022). Explicating AI literacy of employees at digital workplaces. *IEEE Transactions on Engineering Management*, *71*, 810–823. https://doi.org/10.1109/tem.2021.3138503

Chakraborty, R., & Naskar, R. (2024). Role of human physiology and facial biomechanics towards building robust deepfake detectors: A comprehensive survey and analysis. *Computer Science Review*, *54*, 100677. https://doi.org/10.1016/j.cosrev.2024.100677

Cho, J.-H., Cam, H., & Oltramari, A. (2016). Effect of personality traits on trust and risk to phishing vulnerability: Modeling and analysis. In *2016 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)* (pp. 7–13). IEEE. https://doi.org/10.1109/CogSIMA.2016.7497787

Chowdhury, N. H., Adam, M. T., & Teubner, T. (2020). Time pressure in human cybersecurity behavior: Theoretical framework and countermeasures. *Computers & Security*, *97*, 101963. https://doi.org/10.1016/j.cose.2020.101963

CNN. (2024). British engineering giant Arup revealed as >5 million deepfake scam victim. *CNN*. https://edition.cnn.com/2024/05/16/tech/arup-deepfake-scam-loss-hong-kong-intl-hnk/index.html#:~:text=A%20British%20multinational%20design%20and,out%20%2425%20million%20to%20fraudsters.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). Applied multiple regression/correlation analysis for the behavioral sciences (3rd ed.). Routledge. https://doi.org/10.4324/9780203774441

Cole, K. A., Francis, A. L., Rogers, M., & Balazs, J. (2024) Defining Cognitive Workload Through Individual Differences in Capacity-Resource Limitations and Cybersecurity Performance. *Available at SSRN 4918373*.

Creswell, J. W., & Creswell, J. D. (2018). Research Design: Qualitative, Quantitative, and Mixed Methods Approaches (5th ed.). Sage.

Cusack, B., & Adedokun, K. (2018). The impact of personality traits on user's susceptibility to social engineering attacks.

Darwish, A., El Zarka, A., & Aloul, F. (2012, December). Towards understanding phishing victims' profile. In 2012 international conference on computer systems and industrial informatics (pp. 1-5). IEEE.

De Rancourt-Raymond, A., & Smaili, N. (2022). The unethical use of deepfakes. *Journal of Financial Crime*, *30*(4), 1066–1077. https://doi.org/10.1108/jfc-04-2022-0090

Diel, A., Lalgi, T., Schröter, I. C., MacDorman, K. F., Teufel, M., & Bäuerle, A. (2024). Human performance in detecting deepfakes: A systematic review and meta-analysis of 56 papers. *Computers in Human Behavior Reports*, *16*, 100538.

Dolhansky, B., Bitton, J., Pflaum, B., Lu, J., Howes, R., Wang, M., & Ferrer, C. C. (2020). The deepfake detection challenge (dfdc) dataset. arXiv preprint arXiv:2006.07397.

Dsouza, D. S., Hajjar, A. E., & Jahankhani, H. (2024). Deepfakes in social engineering attacks. In *Space Law and Policy* (pp. 153–183). https://doi.org/10.1007/978-3-031-64045-2_8

Devos, H., Gustafson, K., Ahmadnezhad, P., Liao, K., Mahnken, J. D., Brooks, W. M., & Burns, J. M. (2020). Psychometric properties of NASA-TLX and index of cognitive activity as measures of cognitive workload in older adults. *Brain sciences*, *10*(12), 994.

Duran, R., Zavgorodniaia, A., & Sorva, J. (2022). Cognitive load theory in computing education research: A review. *ACM Transactions on Computing Education (TOCE)*, *22*(4), 1-27.

Ei Bolock, & Madany. (2025). Phishing You Understanding Phishing Susceptibility through Personality, Age, Education Level, and Gender. *Medicon Engineering Themes*. https://doi.org/10.55162/mcet.08.263

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical Power Analyses using G*Power 3.1: Tests for Correlation and Regression Analyses. Behavior Research Methods, 41, 1149-1160

Field, A. (2018). Discovering Statistics Using IBM SPSS Statistics (5th ed.). Sage.

Finnish National Board on Research Integrity (TENK). (2019). The ethical principles of research with human participants and ethical review in the human sciences in Finland.

Firc, A., Malinka, K., & Hanáček, P. (2023a). Deepfakes as a threat to speaker and facial recognition: An overview of tools and attack vectors. Heliyon, 9(4), e15090. https://doi.org/10.1016/j.heliyon.2023.e15090

Firc, M., Liz-Lopez, F. J., & Guo, Y. (2023b). Human detection of deepfakes: A cognitive challenge. Computers in Human Behavior, 122, 106835.

Fiss, P. C. (2011). Building better causal theories: A fuzzy set approach to typologies in organization research. *Academy of management journal*, *54*(2), 393-420.

Flick, U. (2018). An Introduction to Qualitative Research (6th ed.). Sage.

Frauenstein, E. D., & Flowerday, S. (2020). Susceptibility to phishing on social network sites: A personality information processing model. Computers & Security, 94, 101862. https://doi.org/10.1016/j.cose.2020.101862

Furnari, S., Crilly, D., Misangyi, V. F., Greckhamer, T., Fiss, P. C., & Aguilera, R. V. (2021). Capturing causal complexity: Heuristics for configurational theorizing. *Academy of Management Review*, *46*(4), 778-799.

Galy, E., Paxion, J., & Berthelon, C. (2017). Measuring mental workload with the NASA-TLX needs to examine each dimension rather than relying on the global score: an example with driving. *Ergonomics*, *61*(4), 517–527. https://doi.org/10.1080/00140139.2017.1369583

Gee, S. L., Höltge, J., Maercker, A., & Thoma, M. V. (2018). Sense of coherence and stress-related resilience: Investigating the mediating and moderating mechanisms in the development of resilience following stress or adversity. Frontiers in psychiatry, 9, 378.

Goh, D. H. (2023). “He looks very real”: Media, knowledge, and search-based strategies for deepfake identification. *Journal of the Association for Information Science and Technology*, *75*(6), 643–654. https://doi.org/10.1002/asi.24867

Goodman, L. A. (2011). Comment: On respondent-driven sampling and snowball sampling in hard-to-reach populations and snowball sampling not in hard-to-reach populations. *Sociological methodology*, *41*(1), 347-353.

Gragg, D. (2003) A Multi-level Defense Against Social Engineering. SANS Institute - as part of Information Security Reading Room

Greavu-Şerban, V., Constantin, F., & Necula, S. C. (2025). Exploring Heuristics and Biases in Cybersecurity: A Factor Analysis of Social Engineering Vulnerabilities. *Systems*, *13*(4), 280

Greckhamer, T., Furnari, S., Fiss, P. C., & Aguilera, R. V. (2018). Studying configurations with qualitative comparative analysis: Best practices in strategy and organization research. *Strategic organization*, *16*(4), 482-495.

Groh, M., Epstein, Z., Firestone, C., & Picard, R. (2021). Deepfake detection by human crowds, machines, and machine-informed crowds. *Proceedings of the National Academy of Sciences*, *119*(1). https://doi.org/10.1073/pnas.2110013119

Guo, Z., Yang, G., Chen, J., & Sun, X. (2021). Fake face detection via adaptive manipulation traces extraction network. Computer Vision and Image Understanding, 204, Article 103170. https://doi.org/10.1016/j.cviu.2021.103170

Halevi, T., Lewis, J., & Memon, N. (2013). A pilot study of cybersecurity and privacy related behavior and personality traits. In Proceedings of the 22nd International Conference on World Wide Web (pp. 737–744). https://doi.org/10.1145/2487788.2488034

Halevi, T., Memon, N., & Nov, O. (2015). Spear-phishing in the wild: A real-world study of personality, phishing self-efficacy and vulnerability to spear-phishing attacks. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.2544742

Halevi, T., Memon, N., Lewis, J., Kumaraguru, P., Arora, S., Dagar, N., et al. (2016). Cultural and psychological factors in cybersecurity. In Proceedings of the 18th International Conference

on Information Integration and Web-based Applications and Services (iiWAS '16) (pp. 318–324). https://doi.org/10.1145/3011141.3011165

Hart, S. G. (2006). NASA-Task Load Index (NASA-TLX); 20 years later. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 50(9), 904–908. https://doi.org/10.1177/154193120605000909

Iannacci, F.& Kraus, S. (2025) Configurational Theory: A review. In S. Papagiannidis (Ed), TheoryHub Book. Available at https://open.ncl.ac.uk / ISBN: 9781739604400

Jenny, G. J., Bauer, G. F., Vinje, H. F., Vogt, K., & Torp, S. (2017). The application of salutogenesis to work. *The handbook of salutogenesis*, 197-210

Kaushal, A., Kumar, S., & Kumar, R. (2024). A review on deepfake generation and detection: bibliometric analysis. Multimedia Tools and Applications, 1-41.

Kapoor, J., & Rahman, N. (2024). Digital Innovation adoption: architectural recommendations and security solutions. In *BENTHAM SCIENCE PUBLISHERS eBooks*. https://doi.org/10.2174/9789815079661124010 1

Kassis, A., & Hengartner, U. (2023, May). Breaking security-critical voice authentication. In 2023 IEEE Symposium on Security and Privacy (SP) (pp. 951–968). IEEE. https://doi.org/10.1109/SP46215.2023.00056

Kaya, B., Abubakar, A. M., Behravesh, E., Yildiz, H., & Mert, I. S. (2020). Antecedents of innovative performance: Findings from PLS-SEM and fuzzy sets (fsQCA). *Journal of Business Research*, *114*, 278-289.

Khan, M. R., Naeem, S., Tariq, U., Dhall, A., Khan, M. N. A., Al Shargie, F., & Al-Nashash, H. (2023, October). Exploring neurophysiological responses to cross-cultural deepfake videos. In *Companion Publication of the 25th International Conference on Multimodal Interaction* (pp. 41-45).

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

Kim, J. H. (2019). Multicollinearity and misleading statistical results. *Korean journal of anesthesiology*, *72*(6), 558-569.

Köbis, N. C., Doležalová, B., & Soraperra, I. (2021). Fooled twice: People cannot detect deepfakes but think they can. *iScience*, *24*(11), 103364. https://doi.org/10.1016/j.isci.2021.103364

Kosch, T., Karolus, J., Zagermann, J., Reiterer, H., Schmidt, A., & Woźniak, P. W. (2023). A survey on Measuring Cognitive Workload in Human-Computer Interaction. *ACM Computing Surveys*, *55*(13s), 1–39. https://doi.org/10.1145/3582272

Korshunov, P., & Marcel, S. (2019, June). Vulnerability assessment and detection of deepfake videos. In 2019 International Conference on Biometrics (ICB) (pp. 1–6). IEEE. https://doi.org/10.1109/ICB45273.2019.8987396

Kulikowski, K., & Ganzach, Y. (2024). The Six Challenges for Personality, Intelligence, Cognitive Skills, and Life Outcomes Research: An Introduction to the Topic. *Journal of Intelligence*, *12*(3), 35.

Lawson, P. A., Crowson, A. D., & Mayhorn, C. B. (2018). Baiting the hook: Exploring the interaction of personality and persuasion tactics in email phishing attacks. In Congress of the International Ergonomics Association (pp. 401–406). Springer. https://doi.org/10.1007/978-3-319-96077-7_42

Levine, T. R. (2014). Truth-default theory (TDT) a theory of human deception and deception detection. *Journal of Language and Social Psychology*, *33*(4), 378-392.

Levine, T. R. (2022). Truth-default theory and the psychology of lying and deception detection. Current Opinion in Psychology, https://doi.org/10.1016/j.copsyc.2022.101380 .

Liz-Lopez, F. J., Firc, M., & Guo, Y. (2024a). The vanishing cues: Human struggles with advanced deepfakes. Forensic Science International: Digital Investigation, 44, 301585. https://doi.org/10.1016/j.fsidi.2023.301585

Liz-Lopez, H., Keita, M., Taleb-Ahmed, A., Hadid, A., Huertas-Tato, J., & Camacho, D. (2024). Generation and detection of manipulated multimodal audiovisual content: Advances, trends and open challenges. Information Fusion, 103, Article 102103. https://doi.org/10.1016/j.inffus.2023.102103

Luca, M., & Zervas, G. (2016). Fake it till you make it: Reputation, Competition, and Yelp Review fraud. *Management Science*, *62*(12), 3412–3427. https://doi.org/10.1287/mnsc.2015.2304

Matli, W., Rancourt-Raymond, A., & Smaili, M. (2024). Psychological impacts of deepfake-enabled fraud on employees. Journal of Business Ethics. https://doi.org/10.1007/s10551-023-05570-2

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of personality*, *60*(2), 175-215.

Mirzaei, A., Carter, S. R., Patanwala, A. E., & Schneider, C. R. (2022). Missing data in surveys: Key concepts, approaches, and applications. *Research in Social and Administrative Pharmacy*, *18*(2), 2308-2316.

Mitnick, K., & Simon, W. L. (2002). The art of deception: Controlling the human element of security. Wiley.

Montañez, R., Golob, E., & Xu, S. (2020). Human cognition through the lens of social engineering cyberattacks. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.01755

Mouton, F., Leenen, L., & Venter, H. S. (2016). Social engineering attack examples, templates and scenarios. Computers & Security, 59, 186–209. https://doi.org/10.1016/j.cose.2016.03.004

Mustak, M., Salminen, J., Mäntymäki, M., Rahman, A., & Dwivedi, Y. K. (2023). Deepfakes: Deceptions, mitigations, and opportunities. Journal of Business Research, 154, 113368.

Olya, H. G., & Altinay, L. (2016). Asymmetric modeling of intention to purchase tourism weather insurance and loyalty. Journal of Business Research, 69(8), 2791-2800.

Ozili, P. K. (2023). The acceptable R-square in empirical modelling for social science research. In *Social research methodology and publishing results: A guide to non-native English speakers* (pp. 134-143). IGI global.

Pappas, I. O., & Woodside, A. G. (2021). Fuzzy-set Qualitative Comparative Analysis (fsQCA): Guidelines for research practice in Information Systems and marketing. *International journal of information management*, *58*, 102310.

Park, J., van den Broek, K. L., Bhullar, N., Ogunbode, C. A., Schermer, J. A., Doran, R., ... & Yadav, R. (2022). Comparison of the inter-item correlations of the Big Five Inventory-10 (BFI-10) between Western and non-Western contexts. *Personality and Individual Differences*, *196*, 111751.

Patel, Y., Tanwar, S., Gupta, R., Bhattacharya, P., Davidson, I. E., Nyameko, R., ... & Vimal, V. (2023). Deepfake generation and detection: Case study and challenges. *IEEE Access*, *11*, 143296-143323.

Pattinson, M., Jerram, C., Parsons, K., McCormac, A., & Butavicius, M. (2012). Why do some people manage phishing e-mails better than others? Information Management & Computer Security, 20(1), 18–28. https://doi.org/10.1108/09685221211219173

Pedersen, K. T., Pepke, L., Stærmose, T., Papaioannou, M., Choudhary, G., & Dragoni, N. (2025). Deepfake-Driven Social Engineering: Threats, Detection Techniques, and Defensive Strategies in Corporate Environments. Journal of Cybersecurity and Privacy, 5(2), 18.

Petty, (1977). ‘‘A cognitive response analysis of the temporal persistence of attitude changes induced by persuasive communications,’’ Ph.D. dissertation, Graduate School, Ohio State Univ., Columbus, OH, USA.

Plass, J. L., Kalyuga, S., & Leutner, D. (2010). Individual differences and cognitive load theory. In *Cognitive Load Theory* (pp. 65-88). Cambridge University Press. https://doi.org/10.1017/CBO9780511844744.006

Price, P. (2015). *Research methods in psychology, 2nd Canadian Edition*. BCcampus.

Qualtrics. (2021). *Qualtrics*. https://www.qualtrics.com/nl/platform/beveiliging/

Ragin, C. C. (2000). Fuzzy-set social science. University of Chicago Press.

Ragin, C. C. (2009). *Redesigning social inquiry: Fuzzy sets and beyond*. University of Chicago press.

Rammstedt, B., & John, O. P. (2006). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, *41*(1), 203–212. https://doi.org/10.1016/j.jrp.2006.02.001

Regulation (EU) 2024/1689, art. 4. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence [2024] OJ L 2024/1689. https://artificialintelligenceact.eu/article/4/

Saunders, M., Lewis, P., & Thornhill, A. (2019). Research Methods for Business Students (8th ed.). Pearson.

Scheeres, Jamison W., "Establishing the Human Firewall: Reducing an Individual's Vulnerability to Social Engineering Attacks" (2008). Theses and Dissertations. 2790. https://scholar.afit.edu/etd/2790

Schneider, C.Q. & Wagemann, C. (2012). Set-Theoretic Methods for the Social Sciences. Cambridge University Press.

Sheng, Z., Fu, J., Jeyaraj, A., & Sun, Y. (2025). Altruistic and egoistic behaviors on enterprise social network platforms: Analysis using PLS-SEM and fsQCA. *Journal of Business Research*, *186*, 114939.

Singh, M., Bhargava, D., Bhargava, A., & Singh, K. (2025). Building resilience: Strategies for business to mitigate deepfake risks. In Deepfakes and Their Impact on Business (pp. 285–298). IGI Global. https://doi.org/10.4018/978-1-6684-8948-2.ch017

Somoray, K., & Miller, D. J. (2023). Providing detection strategies to improve human detection of deepfakes: An experimental study. Computers in Human Behavior, 149, 107917.

Somoray, K., Miller, D., & Holmes (2024), M. Human Performance in Deepfake Detection: A Systematic Review. *Available at SSRN 4955104*.

Sütterlin, S., Lugo, R. G., Ask, T. F., et al. (2022). The role of IT background for metacognitive accuracy, confidence and overestimation of deepfake recognition skills. In Lecture Notes in Computer Science (pp. 103–119). https://doi.org/10.1007/978-3-031-05457-0_9

Tahir, R., Batool, B., Jamshed, H., Jameel, M., Anwar, M., Ahmed, F., ... & Zaffar, M. F. (2021, May). 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).

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International journal of medical education*, *2*, 53.

Guardian. (2024). UK engineering firm Arup falls victim to £20m deepfake scam. *The Guardian*. https://www.theguardian.com/technology/article/2024/may/17/uk-engineering-arup-deepfake-scam-hong-kong-ai-video

Tipton, E., Hallberg, K., Hedges, L. V., & Chan, W. (2017). Implications of small samples for generalization: Adjustments and rules of thumb. *Evaluation review*, *41*(5), 472-505.

Tomlinson. (2022). Learn a new survival skill: Spotting deepfakes. SANS OUCH! Newsletter. https://www.sans.org/newsletters/ouch/learn-a-new-survival-skill-spotting-deepfakes/

Triplett, W. J. (2022). Addressing human factors in cybersecurity leadership. *Journal of Cybersecurity and Privacy*, *2*(3), 573-586.

Uebelacker, S., & Quiel, S. (2014, July). The social engineering personality framework. In *2014 Workshop on Socio-Technical Aspects in Security and Trust* (pp. 24-30). IEEE

Van Der Sloot, B., & Wagensveld, Y. (2022). Deepfakes: regulatory challenges for the synthetic society. *Computer Law & Security Review*, *46*, 105716. https://doi.org/10.1016/j.clsr.2022.105716

Verdoliva, L. (2020). Media forensics and deepfakes: an overview. IEEE journal of selected topics in signal processing, 14(5), 910-932.

Vollrath, M. (2001). Personality and stress. *Scandinavian Journal of Psychology*, *42*(4), 335-347.

Vuorikari, R., Kluzer, S. and Punie, Y., DigComp 2.2, (2022) The Digital Competence framework for citizens; With new examples of knowledge, skills and attitudes, Publications Office of the European Union, https://data.europa.eu/doi/10.2760/115376

Wang, T., Liao, X., Chow, K. P., Lin, X., & Wang, Y. (2024). Deepfake Detection: A Comprehensive Survey from the Reliability Perspective. *ACM Computing Surveys*, *57*(3), 1–35. https://doi.org/10.1145/3699710

Wang, Z., Zhu, H., & Sun, L. (2021). Social engineering in cybersecurity: effect mechanisms, human vulnerabilities and attack methods. *IEEE Access*, *9*, 11895–11910. https://doi.org/10.1109/access.2021.3051633

Wazid, M., Mishra, A. K., Mohd, N., & Das, A. K. (2024). A secure deepfake mitigation framework: architecture, issues, challenges, and societal impact. *Cyber Security and Applications*, *2*, 100040. https://doi.org/10.1016/j.csa.2024.100040

Westerlund, M. (2019). The Emergence of Deepfake Technology: A review. *Technology Innovation Management Review*, *9*(11), 39–52. https://doi.org/10.22215/timreview/1282

Xu, F., Wang, R., Huang, Y., Guo, Q., Ma, L., & Liu, Y. (2022). Countering Malicious DeepFakes: Survey, Battleground, and horizon. *International Journal of Computer Vision*, *130*(7), 1678–1734. https://doi.org/10.1007/s11263-022-01606-8

Yasin, A., Fatima, R., Liu, L., Wang, J., Ali, R., & Wei, Z. (2021). Understanding and deciphering of social engineering attack scenarios. *Security and Privacy*, *4*(4). https://doi.org/10.1002/spy2.161

Yin, R. K. (2018). Case Study Research and Applications: Design and Methods (6th ed.). Sage.

Appendix

A. BFI-10 questionnaire

English version.
Instruction: How well do the following statements describe your personality?

I see myself as someone who …	Disagree strongly	Disagree a little	Neither agree nor disagree	Agree a little	Agree strongly
… is reserved	(1)	(2)	(3)	(4)	(5)
… is generally trusting	(1)	(2)	(3)	(4)	(5)
… tends to be lazy	(1)	(2)	(3)	(4)	(5)
… is relaxed, handles stress well	(1)	(2)	(3)	(4)	(5)
… has few artistic interests	(1)	(2)	(3)	(4)	(5)
… is outgoing, sociable	(1)	(2)	(3)	(4)	(5)
… tends to find fault with others	(1)	(2)	(3)	(4)	(5)
… does a thorough job	(1)	(2)	(3)	(4)	(5)
… gets nervous easily	(1)	(2)	(3)	(4)	(5)
… has an active imagination	(1)	(2)	(3)	(4)	(5)

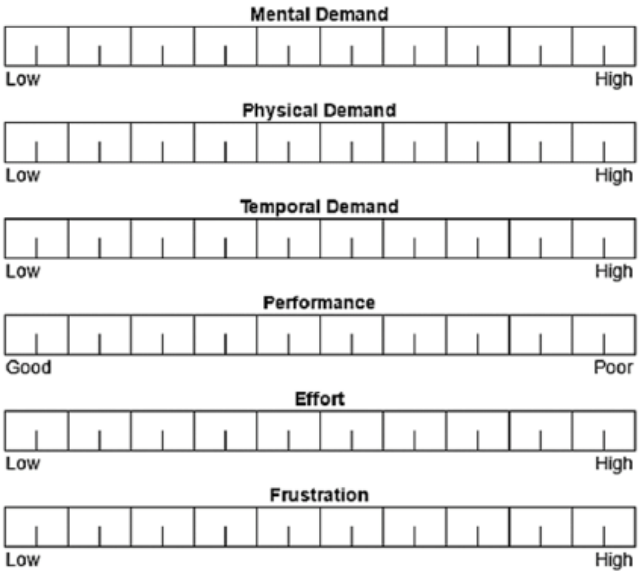
Scoring the BFI-10 scales:
Extraversion: 1R, 6; Agreeableness: 2, 7R; Conscientiousness: 3R, 8; Neuroticism: 4R, 9; Openness: 5R; 10 (R = item is reversed-scored).

Optional additional Agreeableness item (true-scored):

B. SOC-13 questionnaire

Items in the SOC scale: (Likert Scale 1-7)
Q1. Do you have the feeling that you really don't care about what is going on around you? (1-Never, rarely, 2, 3, 4, 5, 6, 7-very often)
Q2. Has it happened in the past that you were surprised by the behaviour of people whom you thought you knew well? (1-Never, 2, 3, 4, 5, 6, 7-always)
Q3. Has it happened that people whom you counted on disappointed you? (1-Never, 2, 3, 4, 5, 6, 7-always)
Q4. Until now your life has had: (1-no clear goals, 2, 3, 4, 5, 6, 7 – very clear goals and purpose)
Q5. Do you have the feeling that you are being treated unfairly? (1- Very often, 2, 3, 4, 5, 6, 7-Very rarely or never)
Q6. Do you have the feeling that you are in an unfamiliar situation and don't know what to do? (1- Very often, 2, 3, 4, 5, 6, 7-Very rarely or never)
Q7. Doing the things you do every day is: (1-a source of deep pleasure and satisfaction, 2, 3, 4, 5, 6, 7 – a source of pain and boredom)
Q8. Do you have very mixed-up feelings and ideas? (1- Very often, 2, 3, 4, 5, 6, 7-Very rarely or never)
Q9. Does it happen that you experience feelings that you would rather not have to endure? (1-Very often, 2, 3, 4, 5, 6, 7-Very rarely or never)
Q10. Many people, even those with a strong character, sometimes feel like losers in certain situations. How often have you felt this way in the past? (1-Never, rarely, 2, 3, 4, 5, 6, 7-very often)
Q11. When certain events occurred, have you generally found that: (1-you overestimated or underestimated their importance, 2, 3, 4, 5, 6, 7 – you assessed the situation correctly?)
Q12. How often do you have the feeling that there is little meaning in the things you do in your daily life? (1- Very often, 2, 3, 4, 5, 6, 7-Very rarely or never)
Q13. How often do you have feelings that you are not sure you can control? (1- Very often, 2, 3, 4, 5, 6, 7-Very rarely or never)
MANAGEABILITY (Behaviour component) (Average punctuation on questions Q3, Q5, Q10, Q13)
COMPREHENSIBILITY (Cognitive component) (Average punctuation on questions Q2, Q6, Q8, Q9, Q11)
MEANINGFULNESS (motivational component) (Average punctuation on questions Q1, Q4, Q7, Q12)
TOTAL SOC (average punctuation on the three aforementioned components)

C. NASA-TLX questionnaire



How much mental and perceptual activity was required (e.g. thinking, deciding, calculating, remembering, looking, searching, etc)? Was the task easy or demanding, simple or complex, exacting or forgiving?

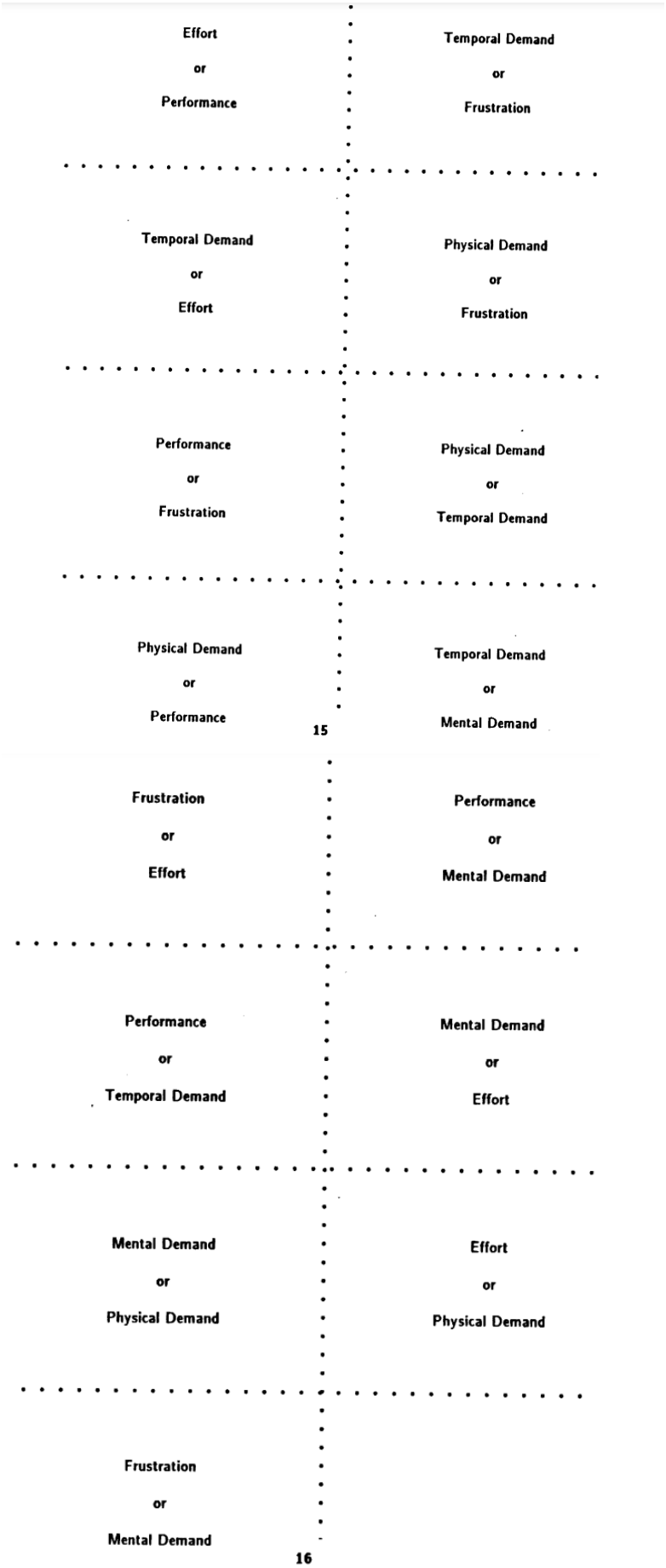
How much physical activity was required (e.g. pushing, pulling, turning, controlling, activating, etc)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?

How much time pressure did you feel due to the rate of pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?

How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?

How hard did you have to work (mentally and physically) to accomplish your level of performance?

How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?


Effort
or
Performance
Temporal Demand
or
Frustration
Temporal Demand
or
Effort
Physical Demand
or
Frustration
Performance
or
Frustration
Physical Demand
or
Temporal Demand
Physical Demand
or
Performance
Temporal Demand
or
Mental Demand
15
Frustration
or
Effort
Performance
or
Mental Demand
Performance
or
Temporal Demand
Mental Demand
or
Effort
Mental Demand
or
Physical Demand
Effort
or
Physical Demand
Frustration
or
Mental Demand
16

D. Participants briefing

1. Welcome to the Survey: Psychological and Cognitive Factors in Deepfake Detection

Thank you for participating in this research study. This survey is part of a master's thesis on the theme: "Psychological and Cognitive Factors in Deepfake Detection."

The study consists of **three short parts**, taking approximately **15 minutes in total**:

2. **Psychological Assessment** (~5 minutes)
   You will be asked to complete a brief set of standardized questions about your personality.

3. **Deepfake Detection Task** (~5 minutes)
   You will view a series of videos and judge whether each one is real or a deepfake.

4. **Post-Test Questionnaire** (~5 minutes)
   After the task, you'll reflect on your experience by answering questions about your perceived workload.

Your Data & Privacy

- **All responses are anonymous** and cannot be traced back to you.
- No personally identifiable information will be collected.
- Data will be handled and stored with **strict confidentiality** and according to ethical research guidelines.
- All data will be stored securely and used **exclusively for academic research** purposes.

If at any time, you feel uncomfortable answering a question, just stop and exit the experimental survey, unfinished records will be deleted automaticly after a week.
Good luck - Sten van Dijk

E. Data Management Plan

i. Research data

Research data refers to all the material with which the analysis and results of the research can be verified and reproduced. It may be, for example, various measurement results, data from surveys or interviews, recordings or videos, notes, software, source codes, biological samples, text samples, or collection data.

In the table below, list all the research data you use in your research. Note that the data may consist of several different types of data, so please remember to list all the different data types. List both digital and physical research data.

Research data type	**Contains personal details/information***	**I will gather/produce the data myself**	**Someone else has gathered/produced the data**	**Other notes**
Example, Data type 1: *Experimental survey*	x	x		

* Personal details/information are all information based on which a person can be identified directly or indirectly, for example by connecting a specific piece of data to another, which makes identification possible. For more information about what data is considered personal go to the Office of the Finnish Data Protection Ombudsman's website

ii. Processing personal data in research

If your data contains personal details/information, you are obliged to comply with the EU's General Data Protection Regulation (GDPR) and the Finnish Data Protection Act. For data that contains personal details, you must prepare a Data Protection Notice for your research participants and determine who is the controller for the research data.

I will prepare a Data Protection Notice** and give it to the research participants before collecting data ☒

The controller** for the personal details is the student themself ☒ the university ☐

My data does not contain any personal data ☐

iii. Permissions and rights related to the use of data

Find out what permissions and rights are involved in the use of the data. Consult your thesis supervisor, if necessary. Describe the use permissions and rights for each data type. You can add more data types to the list, if necessary.

Self-collected data

You may need separate permissions to use the data you collect or produce, both in research and in publishing the results. If you are archiving your data, remember to ask the research participants for the necessary permissions for archiving and further use of the data. Also, find out if the repository/archive you have selected requires written permissions from the participants.

Necessary permissions and how they are acquired:

Experimental Survey Data: The introduction of the experimental survey included clear statements regarding participant anonymity, data storage, and the sensitive nature of psychometric data. Participants were informed that their responses would be treated confidentially and were encouraged to discontinue participation at any point if they felt uncomfortable sharing personal information anonimously. No metadata such as IP addresses was collected or stored; participants were only identified via an anonymous respondent key

iv. Storing the data during the research process

In the university's network drive ☒
In the university-provided Seafile Cloud Service ☐
Other location, please specify: ☒ On Qualtrics – a secure online survey platform affiliated with Tilburg University

Qualtrics is a secure, web-based platform used for creating and distributing online surveys.

Tilburg University provides access to Qualtrics for its students and staff, ensuring data privacy and compliance with institutional standards.

The platform is widely used for academic research due to its robust security features and user-friendly interface (Qualtrics).

The university's data storage services will take care of data security and backup files automatically. If you choose to store your data somewhere other than in the services provided by the university,

please specify how you will ensure data security and file backups. Remember to make sure you know every time where you are saving the edited/modified data.

If you are using a smartphone to record anything, please check in advance where the audio or video will be saved. If you are using commercial cloud services (iCloud, Dropbox, Google Drive, etc.) and your data contains personal data, make sure the information you provide in the Data Protection Notice about data migration matches your device settings. The use of commercial cloud services means the data will be transferred to third countries outside the EU.

v. Documenting the data and metadata

How would you describe your research data so that even an outsider or a person unfamiliar with it will understand what the data is? How would you help yourself recall years later what your data consists of?

Data documentation

Can you describe what has happened to your research data during the research process? Data documentation is essential when you try to track any changes made to the data.
To document the data, I will use:
A field/research journal ☐
A separate document where I will record the main points of the data, such as changes made, phases of analysis, and significance of variables ☐
A readme file linked to the data that describes the main points of the data ☒
Other, please specify: ☐

Data arrangement and integrity

How will you keep your data in order and intact, as well as prevent any accidental changes to it?

I will keep the original data files separate from the data I am using in the research process, so that I can always revert back to the original, if need be. ☒

Version control: I will plan before starting the research how I will name the different data versions and I will adhere to the plan consistently. ☒

I recognise the life span of the data from the beginning of the research and am already prepared for situations, where the data can alter unnoticed, for example while recording, transcribing, downloading, or in data conversions from one file format to another, etc. ☒

Metadata

Metadata is a description of you research data. Based on metadata someone unfamiliar with your data will understand what it consists of. Metadata should include, among others, the file name, location, file size, and information about the producer of the data. Will you require metadata?

I will save my data into an archive or a repository that will take care of the metadata for me. ☒

I will have to create the metadata myself, because the archive/repository where I am uploading the data requires it. ☒

I will not store my data into a public archive/repository, and therefore I will not need to create any metadata. ☒

vi. Data after completing the research

You are responsible for the data even after the research process has ended. Make sure you will handle the data according to the agreements you have made. The university recommends a general retention period of five (5) years, with an exception for medical research data, where the retention period is 15 years. Personal data can only be stored as long as it is necessary. If you have agreed to destroy the data after a set time period, you are responsible for destroying the data, even if you no longer are a student at the university. Likewise, when using the university's online storage services, destroying the data is your responsibility. What happens to your research data, when the research is completed?

I will store all data for 5 years.

If you will store the data, please identify where: In the university drive.

F. Extend truth table analysis High solutions fsQCA

```
--- COMPLEX SOLUTION ---
frequency cutoff: 1
consistency cutoff: 0.803921
                                                      raw       unique
                                                    coverage    coverage   consistency
                                                   ----------  ----------  ----------
~Extra*Agree*Consc*~Neuro*Openn*Soc*~Cworkload     0.179134    0.0450528   0.886873
Extra*Agree*Consc*~Neuro*Openn*Soc*Cworkload       0.329597    0.148523    0.835519
Extra*Agree*~Consc*~Neuro*~Openn*Soc*~Cworkload*~Dexp  0.079543  0.00560462  0.803921
~Extra*Agree*Consc*~Neuro*~Openn*Soc*Cworkload*~Dexp   0.116189  0.0155206   0.815431
Extra*Agree*~Consc*Neuro*Openn*Soc*~Cworkload*~Dexp    0.108213  0.0146583   0.927911
~Extra*~Agree*Consc*Neuro*Openn*Soc*Cworkload*~Dexp    0.119207  0.0394482   0.958406
Extra*Agree*Consc*Neuro*Openn*~Soc*Cworkload*~Dexp     0.13559   0.0105627   0.85462
Extra*Agree*Consc*Neuro*Openn*Soc*~Cworkload*Dexp      0.172235  0.0689805   0.976773
solution coverage: 0.570166
solution consistency: 0.796927


**********************
*TRUTH TABLE ANALYSIS*
**********************

File:  C:/Users/stenv/Desktop/Python/selected_output(correct 2).csv
Model: Daccuracy = f(Extra, Agree, Consc, Neuro, Openn, Soc, Cworkload, Dexp)
Algorithm: Quine-McCluskey

--- PARSIMONIOUS SOLUTION ---
frequency cutoff: 1
consistency cutoff: 0.803921
                        raw       unique
                      coverage    coverage   consistency
                     ----------  ----------  ----------
Neuro                0.533736    0.0691959   0.67817
~Extra*Agree         0.406984    0.0403104   0.716509
Consc*Cworkload      0.630524    0.162751    0.742386
~Consc*~Dexp         0.220522    0           0.704545
~Consc*~Cworkload    0.230222    0           0.810319
solution coverage: 0.841345
solution consistency: 0.627189


**********************
*TRUTH TABLE ANALYSIS*
**********************

File:  C:/Users/stenv/Desktop/Python/selected_output(correct 2).csv
Model: Daccuracy = f(Extra, Agree, Consc, Neuro, Openn, Soc, Cworkload, Dexp)
Algorithm: Quine-McCluskey

--- INTERMEDIATE SOLUTION ---
frequency cutoff: 1
consistency cutoff: 0.803921
Assumptions:
                                                      raw       unique
                                                    coverage    coverage   consistency
                                                   ----------  ----------  ----------
~Extra*Agree*Consc*~Neuro*Openn*Soc*~Cworkload     0.179134    0.0450528   0.886873
Extra*Agree*Consc*~Neuro*Openn*Soc*Cworkload       0.329597    0.148523    0.835519
Extra*Agree*~Consc*~Neuro*~Openn*Soc*~Cworkload*~Dexp  0.079543  0.00560462  0.803921
~Extra*Agree*Consc*~Neuro*~Openn*Soc*Cworkload*~Dexp   0.116189  0.0155206   0.815431
Extra*Agree*~Consc*Neuro*Openn*Soc*~Cworkload*~Dexp    0.108213  0.0146583   0.927911
~Extra*~Agree*Consc*Neuro*Openn*Soc*Cworkload*~Dexp    0.119207  0.0394482   0.958406
Extra*Agree*Consc*Neuro*Openn*~Soc*Cworkload*~Dexp     0.13559   0.0105627   0.85462
Extra*Agree*Consc*Neuro*Openn*Soc*~Cworkload*Dexp      0.172235  0.0689805   0.976773
solution coverage: 0.570166
solution consistency: 0.796927
```

G. Extend truth table analysis Low solutions fsQCA

```
--- COMPLEX SOLUTION ---
frequency cutoff: 1
consistency cutoff: 0.80985
                                                          raw       unique
                                                         coverage  coverage  consistency
                                                         --------  --------  ----------
Agree*Consc*~Neuro*Openn*Soc*~Cworkload*~Dexp            0.300868  0.145975  0.809343
~Extra*~Agree*Consc*~Neuro*~Openn*Soc*~Cworkload*~Dexp   0.114996  0.0206524 0.974155
Extra*Agree*~Consc*~Neuro*~Openn*Soc*~Cworkload*~Dexp    0.106782  0.01502   0.991285
~Extra*Agree*Consc*~Neuro*~Openn*Soc*Cworkload*~Dexp     0.137292  0.0380192 0.885023
Extra*Agree*~Consc*Neuro*Openn*Soc*~Cworkload*~Dexp      0.104436  0.00305092 0.822551
~Extra*~Agree*Consc*Neuro*Openn*Soc*Cworkload*~Dexp      0.113119  0.0260502 0.835355
Extra*Agree*~Consc*~Neuro*~Openn*Soc*Cworkload*Dexp      0.0793241 0.0291012 0.882506
Extra*Agree*Consc*Neuro*Openn*~Soc*Cworkload*~Dexp       0.156771  0.0342643 0.907609
solution coverage: 0.505515
solution consistency: 0.79133


**********************
*TRUTH TABLE ANALYSIS*
**********************

File:  C:/Users/stenv/Desktop/Python/selected_output(correct 2).csv
Model: ~Daccuracy = f(Extra, Agree, Consc, Neuro, Openn, Soc, Cworkload, Dexp)
Algorithm: Quine-McCluskey

--- PARSIMONIOUS SOLUTION ---
frequency cutoff: 1
consistency cutoff: 0.80985
                      raw       unique
                     coverage  coverage  consistency
                     --------  --------  ----------
~Openn               0.512556  0.143863  0.702251
~Cworkload*~Dexp     0.552687  0.136822  0.775436
Neuro*~Dexp          0.386764  0.083783  0.678189
solution coverage: 0.819291
solution consistency: 0.660049


**********************
*TRUTH TABLE ANALYSIS*
**********************

File:  C:/Users/stenv/Desktop/Python/selected_output(correct 2).csv
Model: ~Daccuracy = f(Extra, Agree, Consc, Neuro, Openn, Soc, Cworkload, Dexp)
Algorithm: Quine-McCluskey

--- INTERMEDIATE SOLUTION ---
frequency cutoff: 1
consistency cutoff: 0.80985
Assumptions:
                                                          raw       unique
                                                         coverage  coverage  consistency
                                                         --------  --------  ----------
Agree*Consc*~Neuro*Openn*Soc*~Cworkload*~Dexp            0.300868  0.145975  0.809343
~Extra*~Agree*Consc*~Neuro*~Openn*Soc*~Cworkload*~Dexp   0.114996  0.0206524 0.974155
Extra*Agree*~Consc*~Neuro*~Openn*Soc*~Cworkload*~Dexp    0.106782  0.01502   0.991285
~Extra*Agree*Consc*~Neuro*~Openn*Soc*Cworkload*~Dexp     0.137292  0.0380192 0.885023
Extra*Agree*~Consc*Neuro*Openn*Soc*~Cworkload*~Dexp      0.104436  0.00305092 0.822551
~Extra*~Agree*Consc*Neuro*Openn*Soc*Cworkload*~Dexp      0.113119  0.0260502 0.835355
Extra*Agree*~Consc*~Neuro*~Openn*Soc*Cworkload*Dexp      0.0793241 0.0291012 0.882506
Extra*Agree*Consc*Neuro*Openn*~Soc*Cworkload*~Dexp       0.156771  0.0342643 0.907609
solution coverage: 0.505515
solution consistency: 0.79133
```