Synthetic Impersonation in the Power Sector

Синтетична імперсонація в енергетичному секторі

Karol Jędrasiak [A]

Corresponding author: PhD in Pedagogy, e-mail: kjedrasiak@wsb.edu.pl, ORCID ID: https://orcid.org/0000-0002-2254-1030

Piotr Gawliczek [B]

PhD., Prof., e-mail piotr.gawliczek@uwm.edu.pl, ORCID ID: https://orcid.org/0000-0002-0462-1168

Кароль Єндрасяк [A]

Corresponding author: PhD, e-mail: kjedrasiak@wsb.edu.pl, ORCID ID: https://orcid.org/0000-0002-2254-1030

Пйотр Гавлічек [B]

PhD., професор, e-mail: piotr.gawliczek@uwm.edu.pl, ORCID ID: https://orcid.org/0000-0002-0462-1168

[A] WSB University, Dąbrowa Górnicza, Poland.
[B] University of Warmia and Mazury in Olsztyn, Poland

[A] Академія WSB, м. Домброва-Гурнича, Польща
[B] Вармінсько-Мазурський університет в Ольштині, Польща

Received: July 30, 2026 | Revised: August 28, 2026 | Accepted: August 31, 2026

UDC 004.8:004.056:621.311

DOI: https://doi.org/10.33445/sds.2026.16.4.11

Purpose. This study examines synthetic impersonation in the power sector as a decision-security problem, focusing on the conditions under which manipulated communication may lead to unsafe operational decisions and on the organisational safeguards associated with safer behaviour.

Method. An exploratory, quasi-experimental, scenario-based design was applied. The study comprised 24 attack scenarios, 24 communication stimuli, 8 operational command paths, an 8-layer security control matrix, and a supplier dependency layer. The analytical dataset contained 744 decision observations nested within 31 participants across four procedural conditions. Descriptive and comparative analyses were supplemented with participant-clustered generalized estimating equation modelling.

Findings. Unsafe responses occurred in 26.34% of observations. The unsafe response rate decreased descriptively from 29.03% under Baseline conditions to 23.66% under Full Layered Control, although between-condition differences were not statistically significant. Unsafe responses were concentrated in selected high-consequence scenarios, particularly grid topology disclosure, hot-work approval, and black-start initiation. Unsafe decisions were associated with shorter decision times, whereas perceived authenticity did not meaningfully distinguish safe from unsafe outcomes.

Theoretical implications. The findings support examining synthetic impersonation through a decision-security perspective that complements media-authenticity and deepfake-detection approaches by focusing on the pathway through which communication is translated into operational action.

Practical implications. Power-sector organisations should prioritise high-consequence command paths, strengthen structured verification and authorisation procedures, and separate message receipt from operational execution, including in supplier-mediated communication.

Originality/value. The study provides a sector-specific empirical examination of synthetic impersonation as a human-centred operational security problem and links deepfake research with decision behaviour and critical-infrastructure governance.

Limitations/future research. The exploratory sample and non-significant condition-level effects limit causal and population-level conclusions. Larger, preregistered, multi-organisation studies are required to validate the observed patterns.

Article type: Empirical research article.

Мета роботи. Дослідити синтетичну імперсонацію в енергетичному секторі як проблему безпеки ухвалення рішень, зосереджуючись на умовах, за яких маніпулятивна комунікація може призводити до небезпечних оперативних рішень, а також на організаційних запобіжниках, пов'язаних із безпечнішою поведінкою.

Метод дослідження. Застосовано розвідувальний квазіекспериментальний дизайн на основі сценаріїв. Дослідження охоплювало 24 сценарії атак, 24 комунікаційні стимули, 8 маршрутів оперативних команд, матрицю з 8 рівнів безпекового контролю та шар залежностей від постачальників. Аналітичний набір даних містив 744 спостереження рішень, вкладених у 31 учасника, за чотирьох процедурних умов. Описовий і порівняльний аналіз доповнено моделюванням за допомогою узагальнених оцінювальних рівнянь із кластеризацією за учасниками.

Результати дослідження. Небезпечні відповіді становили 26,34% спостережень. Їхня частка описово зменшилася з 29,03% за базових умов до 23,66% за умов повного багаторівневого контролю, хоча відмінності між умовами не були статистично значущими. Небезпечні відповіді концентрувалися в окремих сценаріях із високими наслідками, зокрема розкриття топології електромережі, погодження вогневих робіт та ініціювання пуску з нульового стану. Небезпечні рішення були пов'язані з коротшим часом ухвалення, тоді як сприймана автентичність суттєво не розрізняла безпечні й небезпечні результати.

Теоретичне значення. Результати підтримують розгляд синтетичної імперсонації крізь призму безпеки ухвалення рішень, яка доповнює підходи до автентичності медіа та виявлення дипфейків, зосереджуючись на маршруті, через який комунікація перетворюється на оперативну дію.

Практичне значення. Організаціям енергетичного сектору слід пріоритезувати маршрути команд із високими наслідками, посилювати структуровані процедури перевірки й авторизації та відокремлювати отримання повідомлення від оперативного виконання, зокрема в комунікації за участю постачальників.

Оригінальність / наукова новизна. Дослідження пропонує галузеве емпіричне вивчення синтетичної імперсонації як людиноцентричної проблеми оперативної безпеки та поєднує дослідження дипфейків із поведінкою під час ухвалення рішень і врядуванням критичної інфраструктури.

Обмеження та перспективи подальших досліджень. Розвідувальна вибірка та статистично незначущі ефекти процедурних умов обмежують причинні висновки й узагальнення на генеральну сукупність. Для перевірки виявлених закономірностей необхідні масштабніші, попередньо зареєстровані дослідження за участю кількох організацій.

Тип статті: Емпірична наукова стаття.

Key words: Synthetic Impersonation; Power Sector; Decision Security; Human Factors; Usable Security; Critical Infrastructure; Social Engineering.

Ключові слова: синтетична імперсонація; енергетичний сектор; безпека ухвалення рішень; людський чинник; зручна у використанні безпека; критична інфраструктура; соціальна інженерія.

Introduction

The rapid diffusion of generative artificial intelligence has fundamentally changed how trusted communication is produced, transmitted, and interpreted in high-consequence environments (Brundage et al., 2018). Synthetic audio, video, and hybrid communication can increasingly reproduce the characteristics of trusted actors and communication contexts. As a result, synthetic impersonation creates a security problem that extends beyond the authenticity of individual media artefacts. In operational environments, the more consequential question is whether manipulated communication can influence a decision, initiate a workflow, bypass an organisational control, or induce unauthorised execution.

This distinction is particularly important in the power sector. Power generation, transmission, distribution, maintenance coordination, outage management, field intervention approval, and emergency response depend on communication-mediated command and authorisation paths. A single unverified instruction may influence switching activity, hot-work approval, restoration logic, maintenance timing, contractor action, or access to sensitive technical information. The consequences of successful impersonation may therefore extend beyond informational or reputational harm to include service disruption, unsafe technical execution, degradation of incident-response capacity, or escalation of an already unstable operational situation.

Operational technology environments further complicate this problem because cybersecurity requirements must coexist with availability, integrity of control, and physical safety. Consequence-oriented guidance for operational technology emphasises that technical and organisational protections should reflect the potential physical effects of disruption, error, and unauthorised control (Stouffer et al., 2023). Similarly, zero trust architecture emphasises explicit verification rather than implicit trust in familiar identities, roles, or communication channels (Rose et al., 2020). These principles become increasingly relevant when synthetic communication can reproduce authority, urgency, and contextual plausibility.

For the purposes of this study, synthetic impersonation is defined as the use of synthetically generated or materially manipulated communication, including audio, video, and hybrid message formats, to impersonate a trusted actor for the purpose of influencing a decision, initiating a workflow, bypassing a control, or inducing unauthorised execution. The power sector provides a particularly relevant environment for examining this phenomenon because operational decisions are frequently time-sensitive, communication takes place through heterogeneous channels, trust relationships extend to external vendors and contractors, and decisions may be made under conditions of incomplete situational awareness or degraded communication quality.

These characteristics create a structural vulnerability when receipt of a plausible message becomes an implicit basis for action. In high-consequence operational environments, the security problem therefore cannot be reduced to whether personnel correctly recognise a manipulated artefact. Even where a communication event appears authentic, organisational design should prevent it from being translated directly into consequential action without appropriate identity verification, authorisation, and procedural control.

The present study addresses this problem through the concept of decision security. Decision security is understood as the organisational ability to approve and execute high-consequence actions only after sufficiently verified identity, valid authorisation, and preserved process integrity. This perspective shifts the analytical focus from the communication artefact itself to the command paths, approval paths, escalation chains, and execution rules through which communication becomes action. The relevant security question is therefore not only whether personnel can identify suspicious communication, but whether organisational procedures prevent an unverified message from progressing into unsafe execution.

Against this background, the main research problem addressed in this article is: under which conditions can synthetic impersonation in the power sector lead to unsafe operational decision making, and which organisational safeguards are most consistent with safer behaviour?

The primary aim of the study is to examine synthetic impersonation as a decision-security problem in the power sector by analysing participant responses to realistic operational communication scenarios under different procedural protection conditions. Rather than evaluating a deepfake detection model in isolation, the study focuses on whether participants move toward execution, verification, escalation, or hold and whether the resulting response is procedurally safe within the operational context presented.

The empirical investigation is structured around three research questions. First, does progressive strengthening of procedural protection reduce the unsafe response burden under synthetic impersonation conditions in the power sector? Second, is unsafe decision behaviour concentrated in selected high-consequence scenarios rather than distributed uniformly across the communication set? Third, are unsafe outcomes better explained by behavioural compression and weak procedural routing than by perceived artefact authenticity alone?

Based on the decision-security framework, three expected empirical patterns were formulated prior to the interpretation of the results. First, unsafe responses were expected to decrease directionally as control logic became more layered. Second, unsafe outcomes were expected to concentrate in scenarios involving technically consequential approval, disclosure, restoration, or interruption actions. Third, perceived authenticity was expected to provide relatively weak separation between safe and unsafe outcomes, whereas unsafe responses were expected to be associated with faster decision behaviour.

To examine these expectations, the study employs an exploratory, quasi-experimental, scenario-based design. Participant responses are analysed across realistic synthetic impersonation scenarios involving operational, maintenance, emergency, and supplier-related decision paths and across progressively strengthened procedural conditions. The analysis combines behavioural outcomes, decision time, perceived authenticity, procedural response categories, and organisational control structures. Given the exploratory character of the study and the participant sample, the empirical expectations are treated as directional rather than as a basis for broad population-level causal claims.

The contribution of the study is threefold. First, it provides a sector-specific empirical extension of the broader decision-security proposition that synthetic impersonation in high-trust environments should be addressed through protected decision pathways and safeguarded execution logic rather than through media inspection alone (Jędrasiak, 2026). Second, it operationalises this proposition through realistic power-sector command and authorisation scenarios. Third, it examines whether layered procedural controls, including callback verification, second-person confirmation, execution hold, and broader layered protection, are associated with safer decision behaviour. In this way, the article connects synthetic-media research with human-centred cybersecurity, safety-oriented systems thinking, and operational governance in critical infrastructure.

Literature Review

Research relevant to synthetic impersonation spans several partially overlapping fields, including deepfake generation and detection, social and informational effects of synthetic media, social engineering, usable security, and cybersecurity in critical infrastructure. Bringing these research streams together is necessary because synthetic impersonation in operational environments combines properties of a manipulated communication artefact with authority cues, organisational trust, time pressure, and established decision procedures.

A substantial part of the deepfake literature has focused on generation, detection, and media forensics. Verdoliva (2020) examines the development of media-forensic approaches in response to increasingly sophisticated manipulation techniques. Tolosana et al. (2020) provide a broader review of face manipulation and fake-detection methods, while Mirsky and Lee (2021) analyse both the creation and detection of deepfakes. These studies demonstrate the technological complexity of identifying synthetic content and establish media authenticity as an important component of the wider security problem.

Research has also extended technical analysis across different communication modalities. Jędrasiak (2024b) examines audio-stream analysis for deepfake threat identification, while Jędrasiak and Wolański (2023) investigate audio-video consistency in the detection of manipulated public-speaking content. Image-processing artefacts and deep-network representations have similarly been examined as indicators of modified or synthetic content (Jędrasiak, 2025; Jędrasiak & Bijoch, 2025). These approaches remain important for identifying potentially manipulated artefacts, particularly as synthetic communication becomes increasingly multimodal.

Technical detection, however, represents only one dimension of the problem. A second strand of research examines the effects of deepfakes on deception, uncertainty, and trust. Vaccari and Chadwick (2020) show that synthetic political video can influence uncertainty and trust in information environments, while Hancock and Bailenson (2021) emphasise the broader social consequences of deepfakes. Kietzmann et al. (2020) and Westerlund (2019) likewise conceptualise deepfake technology as a challenge with organisational and societal implications extending beyond the manipulated artefact itself.

The limitations of relying exclusively on human recognition are particularly relevant in this context. Experimental evidence presented by Köbis et al. (2021) indicates that individuals may have difficulty detecting deepfakes while simultaneously remaining overconfident in their ability to do so. Consequently, the operational resilience of an organisation cannot depend solely on an individual's capacity to determine whether a communication artefact is genuine. This problem becomes particularly significant when the communication concerns an action that is time-sensitive, technically consequential, or associated with a trusted authority figure.

Synthetic impersonation also intersects directly with the literature on social engineering. Workman (2008) demonstrates that phishing and pretexting exploit cognitive and behavioural mechanisms and can use authority and contextual plausibility to influence security-related behaviour. Reviews of contemporary phishing training indicate that awareness interventions continue to face difficulties in generating durable protective behaviour (Marshall et al., 2024). Aldawood and Skinner (2019) similarly identify weaknesses in social-engineering awareness programmes when protective behaviour is insufficiently connected to practical organisational contexts. More recent work on deepfake-driven social engineering suggests that synthetic media may further increase the plausibility and scalability of human-targeted attacks and that effective defence must combine technical detection with procedural and organisational safeguards (Pedersen et al., 2025).

These findings are consistent with the broader literature on usable and human-centred security. Adams and Sasse (1999) argue that security failures cannot simply be attributed to users when security mechanisms themselves conflict with practical working behaviour. Beautement et al. (2008) further demonstrate that security compliance imposes costs on users and that excessive procedural burden can influence whether protective requirements are followed. Egelman and Peer (2015) similarly emphasise behavioural dimensions of security practice. Taken together, this literature suggests that successful protective behaviour depends not only on threat awareness but also on whether security procedures remain usable under workload, time pressure, and competing operational requirements.

This perspective is particularly relevant to critical infrastructure. Ghafir et al. (2018) identify the human factor as an important component of critical-infrastructure vulnerability, while Knowles et al. (2015) demonstrate that cybersecurity management in industrial control systems requires consideration of organisational, procedural, and technical factors. Cherdantseva et al. (2016) likewise show that cyber-risk assessment in supervisory control and data acquisition environments must account for the specific characteristics of cyber-physical systems. In such systems, cybersecurity cannot be separated from operational continuity and physical consequences.

Operational technology guidance reinforces this consequence-oriented approach. Stouffer et al. (2023) emphasise that OT cybersecurity measures should account for safety, reliability, availability, and potential physical effects. Zero trust architecture similarly rejects implicit trust based solely on network position or assumed identity and instead emphasises explicit verification and policy-governed access (Rose et al., 2020). Incident-handling guidance additionally highlights the importance of structured verification, escalation, evidence preservation, and controlled response (Cichonski et al., 2012). Applied to synthetic impersonation, these principles imply that apparent sender identity or familiarity of a communication channel should not independently provide sufficient authorisation for high-consequence execution.

The power sector adds a further layer of complexity because operational trust frequently extends beyond the internal organisational boundary. Vendors, original equipment manufacturers, maintenance contractors, integrators, and technical specialists may participate in access requests, maintenance coordination, troubleshooting, restoration, or other technically consequential processes. As a result, the effective security boundary includes supplier-mediated communication and escalation chains. Informal callback practices, incomplete escalation arrangements, or excessive reliance on familiar external contacts may therefore create pathways through which synthetic impersonation can influence operational decisions.

The concept of decision security provides a framework for integrating these research streams. Jędrasiak (2026) argues that in high-trust environments the primary security asset is not only the authenticity of the communication artefact but also the decision pathway that the artefact attempts to influence. Under this perspective, communication receipt, identity verification, authorisation, escalation, execution, and logging constitute distinct stages that should not collapse into a single act of trust. Synthetic impersonation becomes operationally consequential when manipulated communication can bypass these stages and move directly from message receipt to action.

This perspective does not diminish the importance of deepfake detection. Instead, it locates detection within a broader security architecture. Media-forensic analysis may contribute evidence regarding the authenticity of a communication artefact, but it does not independently determine whether a requested operational action is authorised, safe, or procedurally valid. In high-consequence environments, artefact assessment and execution control therefore represent complementary rather than interchangeable security functions.

Despite substantial research on deepfake detection, social engineering, human-centred cybersecurity, and operational technology security, an empirical gap remains at the intersection of these fields. Existing studies provide strong foundations for understanding synthetic-media authenticity, human susceptibility to deception, security behaviour, and organisational controls. Comparatively less empirical attention has been devoted to how synthetic impersonation affects decision behaviour within high-consequence operational command paths and whether unsafe responses are associated primarily with perceived artefact authenticity or with behavioural and procedural factors occurring between communication receipt and action.

The present study addresses this gap within the power sector. It examines participant behaviour across realistic operational scenarios and different procedural protection conditions, with particular attention to unsafe action selection, verification and escalation behaviour, decision time, perceived authenticity, and the concentration of unsafe outcomes within specific high-consequence

command paths. By doing so, the study links technical deepfake research with human-centred security and operational governance and empirically examines whether protecting the pathway from communication to execution provides a useful framework for understanding synthetic impersonation risk.

Materials and Methods

The study was designed as an exploratory, quasi-experimental, scenario-based investigation of synthetic impersonation in the power sector. Its primary aim was to assess how participants responded to communication events that simulated realistic impersonation attempts directed at operational, maintenance, emergency, and supplier-related decision paths. The study did not attempt to measure perceptual detection performance alone. Instead, it focused on decision behaviour and procedural safety, treating unsafe execution or incorrect routing as the principal outcome of interest. The design combined several structured data layers. These included an operational path inventory, an attack scenario library, a stimulus metadata layer, a participant response dataset, a security control matrix, and a supplier dependency dataset. Together, these layers provided a decision-oriented empirical framework in which communication artefacts, organisational controls, and participant responses could be analysed jointly rather than in isolation.

An operational path inventory was developed to identify communication dependent actions in the power sector that should be treated as protected decision pathways. The inventory comprised 8 operational paths representing technically or organisationally consequential actions. These paths included operational switching, emergency work interruption, maintenance override situations, and other communication linked authorisation contexts. For each path, the dataset recorded action name, action type, typical communication channel, sender role, time pressure, safety relevance, current verification practice, and criticality. This layer was necessary because the study was not intended to analyse synthetic communication in abstract terms. Each scenario had to be embedded in a concrete operational context in which receipt of a message could plausibly affect technical execution, continuity, safety, or escalation logic.

A scenario library containing 24 attack scenarios was constructed. Each scenario represented a plausible synthetic impersonation attempt relevant to the power sector. The scenarios covered high-consequence goals such as load shedding initiation, virtual private network access approval, maintenance cancellation, fault clearing, outage approval, work authorisation, grid topology disclosure, black start initiation, and hot work approval. Each scenario included at least six core attributes: scenario identifier, attack goal, claimed sender role, communication modality, urgency class, and the safe response expected under secure operating conditions. The sender roles included director, dispatcher, vendor engineer, and other organisational or supplier linked actors. The modality layer included audio, audio with noise, voicemail, hybrid communication, and video deepfake. The urgency layer covered low, medium, high, critical, and extreme situations. The scenario set was intentionally constructed to reflect not only media manipulation, but also contextual plausibility, authority cues, supplier involvement, and operational pressure. This was important because the study was designed to examine whether synthetic impersonation succeeds through media realism alone or through the interaction between signal, context, and decision process.

Each scenario was linked to one stimulus entry in the stimuli metadata layer, yielding 24 stimuli in total. Of these, 18 were audio files in Waveform Audio File Format (WAV) and 6 were video files in MPEG-4 (MP4) containers. Eighteen stimuli were marked as synthetic and 6 as reference or non-synthetic. Noise or channel degradation was intentionally varied and included low, low-studio, medium, medium-room, high-GSM, high-wind, and high-machinery conditions. Claimed sender identities covered both internal and external roles, including director, dispatcher, vendor expert, security operations lead, health-safety-environment lead, and chief technology officer. The stimulus documentation recorded the communication format, scenario linkage, sender role, modality, and channel condition for each item.

The analysis was therefore framed at the level of behavioural decision response to scenario-valid communication artefacts, rather than as a comparative assessment of individual synthesis technologies or media-forensic detectors. The inclusion of degraded audio conditions was deliberate because communication in field operations and emergency coordination often occurs under noisy, compressed, or imperfect transmission conditions.

The main analytical dataset contained 744 decision observations generated by 31 participants. Participants were grouped into four operationally relevant role categories: Safety Officer, Operator, Dispatcher, and Maintenance Engineer. Years of experience were recorded for each participant. The participant pool should be interpreted as an exploratory operational sample rather than as a statistically representative cross section of the entire power sector workforce. Each observation represented one participant response to one scenario linked communication event under one procedural condition. The dataset included participant identifier, role group, years of experience, scenario identifier, stimulus identifier, condition, decision time in seconds, actual decision, confidence, authenticity score, urgency perception, and response label. The response label served as the primary safety outcome variable and indicated whether the participant response was procedurally unsafe in the context of the scenario.

The design included four procedural conditions, each represented by 186 observations: Baseline, Callback Only, Callback plus Second Person, and Full Layered Control. The Baseline condition represented the absence of strengthened procedural safeguards beyond the participant's default response logic. Callback Only introduced an explicit requirement for out of band verification before action. Callback plus Second Person added an additional confirmation requirement for relevant decisions. Full Layered Control represented the most mature protective condition and combined stronger verification, confirmation, and broader procedural safeguards consistent with a layered resilience approach. The condition structure was designed to evaluate whether progressively stronger decision security measures were associated with lower unsafe response burden. In this sense, the study operationalised organisational protection not as a binary presence or absence of detection, but as an increasingly structured pathway from message receipt to authorised execution.

An 8-layer security matrix was used to document the organisational control architecture relevant to protection of decision pathways. The layers included Channel, Identity, Threshold, Execution, Supplier, Incident, Physical, and Public Communications. Each control entry contained the control name, activation trigger, and expected evidentiary trace. The purpose of this layer was not merely descriptive. It provided a formal structure for mapping each scenario to the class of protection that should have interrupted unsafe progression. This allowed interpretation of participant outcomes in relation to control design rather than in relation to message content alone.

Because the power sector frequently depends on external vendors, original equipment manufacturers, maintenance contractors, and integrators, a supplier dependency dataset was included. This layer comprised 5 supplier relationship types and recorded whether the supplier could initiate command related communication, whether a named escalation contact existed, whether callback procedures were mandatory or informal, and which notification clauses applied. This dataset served two methodological purposes. First, it extended the analysis beyond the internal organisational perimeter. Second, it allowed the study to examine whether communication dependent trust chains involving suppliers and contractors represented a structurally relevant part of the synthetic impersonation risk surface.

The primary outcome variable was the binary response label indicating whether the participant's response was unsafe for the given scenario. A value of one denoted an unsafe response, while zero denoted a safe or procedurally appropriate response. Secondary behavioural variables included the participant's actual decision category, recorded as Execution, Verification, Escalation, or Hold, together with decision time in seconds. Perceptual variables included

confidence, authenticity score, and urgency perception. These variables were retained in order to test whether unsafe responses were more strongly associated with subjective authenticity judgement or with behavioural compression and decision process characteristics.

The analysis was primarily descriptive and comparative, with exploratory inferential testing. Unsafe response burden was summarised across conditions, response categories, roles, scenario families, modalities, and urgency classes. Cross-tabulations were evaluated using chi-square tests of independence. Continuous variables such as decision time, confidence, authenticity score, and years of experience were compared between safe and unsafe outcomes using Welch's t-test. Because 744 observations were nested within 31 participants, the observation-level analyses were supplemented with a participant-clustered generalized estimating equation logistic model including condition, role group, authenticity score, and decision time. This additional model was used to distinguish the most robust behavioural signal from descriptive between-group patterns. Statistical significance was interpreted at alpha=0.05, but inferential findings were treated as exploratory due to the modest participant sample and multi-factor design. Given the scenario-based structure, the repeated observations nested within participants, and the moderate participant sample, inferential results were treated conservatively. The principal analytical objective was not to make broad population claims, but to assess whether the observed response pattern was more consistent with a media-authenticity interpretation or with a decision-security interpretation centred on behavioural timing and protected routing.

All structured data layers were prepared in tabular form and stored in Comma Separated Values and Microsoft Excel Workbook formats. The participant response dataset, scenario library, operational path inventory, stimuli metadata, security matrix, and supplier dependency layer were integrated through common identifiers, allowing direct linking of responses to scenarios, stimuli, and control requirements. This structure supports reproducible descriptive analysis and allows later expansion through additional participants, scenario families, or sector-specific subclasses of command paths. Several design features were intended to strengthen interpretive validity. First, all four procedural conditions were balanced in terms of the number of observations, which reduced the risk that the main condition comparison was driven by unequal exposure volume. Second, each scenario was defined together with an expected safe response pathway, making it possible to evaluate participant behaviour against a scenario specific procedural benchmark rather than against a generic intuition-based notion of suspicion. Third, the study integrated operational path, scenario, stimulus, supplier, and control layers through shared identifiers, which improved traceability between communication context, expected safeguard, and observed response. At the same time, the study was not designed to estimate real-world incident prevalence or to provide definitive causal evidence regarding the superiority of one procedural configuration over another. No priori power target was set for condition-level effects, and the multi-factor structure of the experiment means that between-condition contrasts should be interpreted as hypothesis-generating rather than conclusive. A larger, preregistered, multi-organisation replication would be necessary for strong comparative claims about layered controls. The main components of the study design and the composition of the analytical dataset are summarised in Table 1.

Table 1: Study design and dataset composition

Component	Value
Participants	31
Decision observations	744
Operational roles	4
Operational command paths	8
Attack scenarios	24
Stimuli	24
Procedural conditions	4

Component	Value
Security control layers	8
Supplier relationship types	5

Source: *compiled by the authors based on the study dataset.*

Results

The final analytical dataset comprised 744 decision observations generated by 31 participants exposed to 24 synthetic impersonation scenarios and 24 associated stimuli. The participants represented four role groups relevant to power sector operations: Safety Officers, Operators, Dispatchers, and Maintenance Engineers. The study design included four balanced procedural conditions, namely Baseline, Callback Only, Callback plus Second Person, and Full Layered Control, each represented by 186 observations. In addition to the response dataset, the empirical package included an inventory of 8 operational command paths, a scenario library of 24 attack situations, a stimulus metadata layer, an 8 layer security control matrix, and a supplier dependency layer covering 5 external relationship types. This structure allowed the analysis to move beyond binary deepfake detection logic and examine the operational consequences of synthetic impersonation in a decision-oriented framework. Such a perspective is consistent with prior work arguing that high-consequence sectors should treat decision pathways, approval chains, and operational command logic as security relevant assets rather than focusing exclusively on artefact classification.

Across the full dataset, 196 of 744 observations were labelled as unsafe, corresponding to an overall unsafe response rate of 26.34%. This means that slightly more than one quarter of all tested decision events resulted in a response pattern that was procedurally inconsistent with the required safe path for the given scenario. Conversely, 548 observations, or 73.66%, were classified as safe. This overall level indicates that synthetic impersonation can materially affect operational decision making in the power sector while preserving sufficient variability to compare procedural safeguards, role groups, modalities, and scenario families.

The behavioural distribution was relatively balanced across the four response categories. There were 198 Execution responses, 197 Verification responses, 189 Escalation responses, and 160 Hold responses. This is analytically valuable because it indicates that participants did not collapse into a single dominant behavioural pattern. Instead, they used a spectrum of decision strategies ranging from immediate action to procedural delay or escalation. When unsafe response burden was analysed by response category, Verification yielded the lowest unsafe proportion at 21.83%. Hold produced an unsafe proportion of 25.00%, Execution 27.78%, and Escalation the highest at 30.69%. The global association between response category and safety outcome did not reach statistical significance, chi-square(3)=4.27, p=0.234, Cramer's V=0.076. Although these differences should therefore be interpreted cautiously, the pattern is operationally important. It suggests that controlled verification was the most reliable immediate response strategy in this dataset, whereas escalation was not automatically safe. In practice, escalation may remain unsafe if it proceeds through the wrong actor, the wrong channel, or without maintaining execution control.

The central empirical question concerned whether progressively stronger organisational controls were associated with lower unsafe response rates. The observed pattern was directionally consistent with this expectation. In the Baseline condition, the unsafe response rate was 29.03% with 54 unsafe observations out of 186. In the Callback Only condition, it declined to 26.34% with 49 unsafe observations. The Callback plus Second Person condition produced the same unsafe rate of 26.34%, again with 49 unsafe observations. The Full Layered Control condition showed the lowest unsafe response rate at 23.66%, corresponding to 44 unsafe observations. Thus, the total reduction from Baseline to Full Layered Control was 5.37 percentage points, equivalent to a relative reduction of approximately 18.5% in unsafe responses. This descriptive gradient is compatible with the hypothesised direction of effect, but it cannot support strong inferential claims. A chi-square test

comparing unsafe and safe outcomes across the four conditions was non-significant, chi-square(3)=1.39, p=0.709, Cramer's V=0.043. The condition contrasts should therefore be treated as exploratory patterns rather than as evidence that one procedural configuration is operationally superior. The distribution of unsafe responses across the four procedural conditions is presented in Table 2.

Table 2: Unsafe response burden by procedural condition

Condition	Unsafe, n	Total, n	Unsafe rate
Baseline	54	186	29.03%
Callback Only	49	186	26.34%
Callback plus Second Person	49	186	26.34%
Full Layered Control	44	186	23.66%

Source: *calculated by the authors based on the study dataset.*

Unsafe response burden differed across operational roles. Maintenance Engineers showed the highest unsafe response rate at 31.94%, followed by Dispatchers at 27.08%, Operators at 25.00%, and Safety Officers at 23.61%. In absolute terms, this corresponded to 46 unsafe observations among Maintenance Engineers, 39 among Dispatchers, 60 among Operators, and 51 among Safety Officers. These differences again should be interpreted conservatively, because the overall association between role group and unsafe response label was not statistically significant, chi-square (3) = 3.42, p = 0.331, Cramer's V = 0.068. Nevertheless, the descriptive pattern is meaningful for engineering and governance interpretation. It suggests that personnel involved in maintenance and technically consequential coordination may require stronger procedural reinforcement, especially in contexts where requests appear routine, technically plausible, and time sensitive.

Unsafe outcomes were not evenly distributed across scenarios. The highest unsafe response rate was recorded for SCEN_09, titled Share Grid Topology, at 41.94%. This was followed by SCEN_21, Approve Hot Work, at 38.71%, and SCEN_18, Trigger Black Start, at 35.48%. A second cluster of high burden scenarios included Log Request, Halt Turbine, Fault Clearing, and Reroute Power, each at 32.26%. This concentration is one of the most important empirical findings in the dataset. It demonstrates that the risk of unsafe decision making under synthetic impersonation is not uniformly distributed across communication events. Instead, vulnerability accumulates around specific high-consequence contexts involving sensitive technical knowledge, work approval, restoration logic, or operational reconfiguration. This directly supports the decision pathway perspective, which treats operational command paths as security assets and argues that protection should be concentrated at high-consequence decision points. The scenarios with the highest observed unsafe response rates are presented in Table 3.

Table 3: Scenarios with the highest unsafe response burden.

Scenario ID	Scenario title	Unsafe rate
SCEN_09	Share Grid Topology	41.94%
SCEN_21	Approve Hot Work	38.71%
SCEN_18	Trigger Black Start	35.48%
SCEN_04	Log Request	32.26%
SCEN_06	Fault Clearing	32.26%
SCEN_11	Halt Turbine	32.26%
SCEN_19	Reroute Power	32.26%

Source: calculated by the authors based on the study dataset.

The data also showed variation by communication modality. The highest unsafe response rate was observed for Audio plus Noise scenarios at 33.06%. Standard Audio scenarios produced 25.40%, Hybrid scenarios 25.00%, Video Deepfake scenarios 25.00%, and Voicemail scenarios 24.19%. Although the overall modality association was not statistically significant at the global level, chi-square (4) = 3.53, p = 0.474, Cramer's V=0.069, the descriptive pattern remains highly relevant from an operational perspective. The fact that noisy audio produced the highest unsafe response burden suggests that degraded communication may in some cases be more dangerous than visually compelling synthetic video. In operational settings, ambiguity may interact with urgency and role authority in a way that reduces reflective verification and increases procedural compliance.

The highest unsafe response rates were recorded for High urgency and Critical urgency scenarios, both at 29.03%. Medium urgency scenarios produced 26.34%, whereas Low urgency and Extreme urgency scenarios each produced 23.66%. The global association between urgency category and unsafe response label was not statistically significant, with p = 0.721. The absence of a monotonic urgency gradient does not mean that urgency is irrelevant. Instead, it suggests that urgency functions in combination with other contextual variables such as sender role, modality, expected action, and available safeguard. This is compatible with the analytical logic of the decision security perspective, which argues that operational error arises not from any single signal alone, but from the interaction of contextual plausibility, authority cues, signal characteristics, and limited time for challenge and verification. The overall urgency comparison was not statistically significant, chi-square (4) = 2.08, p=0.721, Cramer's V = 0.053.

One of the most notable findings concerns the weak separation between perceived authenticity and actual safety outcome. The mean authenticity score was 3.08 for safe responses and 3.05 for unsafe responses. This difference was negligible and statistically non-significant, p=0.756. Confidence also did not differ meaningfully between safe and unsafe responses, with mean values of 3.01 and 3.04 respectively, p = 0.768. By contrast, decision time showed a clearer behavioural signal. Unsafe responses were faster, with a mean response time of 74.73 seconds compared with 82.26 seconds for safe responses. This difference was statistically significant, p=0.025. Experience showed only a weak descriptive pattern, with unsafe responses associated with slightly lower mean experience, 17.02 years versus 18.31 years, and this effect did not reach significance, p=0.062. This pattern is illustrated in Fig. 1.

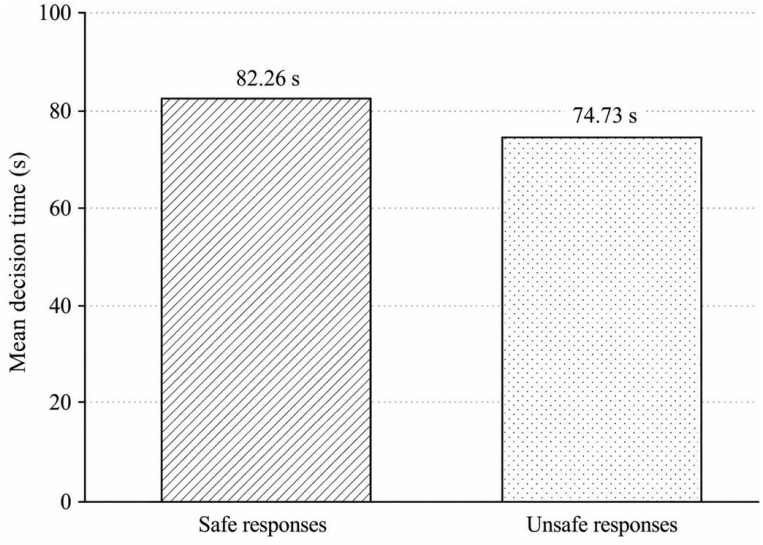

Figure 1: Mean decision time for safe and unsafe responses. Unsafe responses were faster on average (74.73 s) than safe responses (82.26 s). This visualisation corresponds to the Welch's t-test reported in the text (p = 0.025).

Source: calculated by the authors based on the study dataset.

To account for repeated observations nested within participants, a participant-clustered generalized estimating equation logistic model was estimated. In this model, decision time remained significantly associated with lower odds of an unsafe response (beta=-0.0045, z=-2.26, p=0.024), whereas authenticity score did not (beta=-0.0257, z=-0.38, p=0.707). None of the condition contrasts reached statistical significance in the clustered model (all p>=0.251). Substantively, each additional 10 seconds of decision time was associated with an estimated approximately 4.4% reduction in the odds of an unsafe response. This model reinforces the interpretation that behavioural compression, rather than perceived realism, was the strongest signal in the present dataset.

The scenario library and supplier dependency layer also support a more focused interpretation of trust extension beyond the internal perimeter. Several high burden scenarios involved supplier related, contractor related, or externally mediated authorisation logic, including VPN Access, Approve Outage, and Trigger Black Start. The supplier dependency dataset further showed that some external relationships relied on informal callback procedures or incomplete escalation arrangements. This pattern is directly aligned with the supplier assurance argument advanced in earlier decision security work, which states that decision security extends beyond the organisation's internal boundary and that weaknesses in external communication, logging, notification, and escalation arrangements can create equally serious paths for manipulation.

Discussion

The present findings provide provisional evidence consistent with the view that synthetic impersonation in the power sector is better analysed as a decision security problem than as a narrow media-authenticity problem. The strongest empirical signal in the dataset did not concern realism judgments. It concerned the point at which a plausible communication event was translated into action under time pressure. This result is strongly consistent with the conceptual structure of the broader decision security model. That framework argues that the decisive object of protection is not the media object itself, but the decision pathway that the artefact attempts to influence. The empirical material presented here provides a sector-specific extension of that proposition in a power sector context.

A particularly important result is the near absence of difference in authenticity perception between safe and unsafe outcomes. If unsafe decisions were primarily caused by a stronger subjective belief that a message was genuine, authenticity scores should have separated the two groups more clearly. They did not. Instead, the behavioural difference that did emerge, both in simple comparisons and in the participant-clustered model, was response speed. This is methodologically and practically significant. It suggests that decision failure under synthetic impersonation cannot be reduced to perceptual deception alone. The broader usable security literature has long argued that security failures often emerge from friction, workload, and procedural misalignment rather than from a simple lack of threat awareness (Adams & Sasse, 1999; Beautement et al., 2008; Egelman & Peer, 2015). The present dataset is consistent with that view. In other words, the problem is not merely that the recipient believes the artefact. The deeper problem is that the recipient commits too early, before routing the event through a protected verification and authorisation process.

The pattern observed across procedural conditions is also important, but it must be interpreted cautiously. Unsafe response rates decreased gradually from Baseline to Full Layered Control, yet these differences were not statistically significant either in the global observation-level comparison or in the participant-clustered model. Accordingly, the present study does not provide conclusive evidence that one tested procedural regime outperformed the others. Instead, it shows that the descriptive pattern is compatible with the hypothesised direction of effect and therefore warrants larger replication. This exploratory pattern remains theoretically relevant because it is

compatible with, but does not prove, the layered-resilience logic proposed in earlier decision-security work. The current dataset therefore supports a restrained claim: layered controls remain a plausible design hypothesis, but the operational superiority of any particular control bundle should not be inferred from the present sample alone.

The scenario specific findings strengthen the claim that operational command paths should be treated as explicit security assets. The highest unsafe response rates were associated with scenarios involving grid topology sharing, hot work approval, black start initiation, turbine halt, and rerouting of power. These are not generic communication events. They are points at which communication intersects directly with authorisation, technical state change, restoration logic, or work coordination. This concentration effect matters because it provides a practical basis for targeted hardening. The results imply that power sector organisations should not try to impose maximum friction across all communications. Instead, they should map and prioritise those command paths in which a single unverified instruction could alter technical state, disrupt continuity, or degrade incident response. Such an approach follows the reasoning of the decision security model, which explicitly links synthetic impersonation risk to high impact decision categories and argues for stronger safeguards only where consequences justify them.

The higher unsafe burden observed among Maintenance Engineers, together with the scenario burden linked to supplier relevant actions, points toward a structurally important interpretation. In the power sector, maintenance, contractor coordination, and supplier-mediated access often occupy a grey zone between routine technical cooperation and high-consequence decision making. That grey zone may encourage informal trust assumptions, especially when the communication content appears technically plausible and operationally ordinary. This is where supplier assurance becomes more than a procurement matter. It becomes a decision security issue. The empirical material suggests that communication dependent trust chains involving vendors, integrators, contractors, or external specialists deserve stronger escalation design, documented callback procedures, named contacts, and evidentiary logging. This interpretation is directly aligned with the position that supplier assurance forms part of the effective security boundary rather than a secondary administrative layer.

From an engineering and governance perspective, the present findings are best treated as exploratory design implications rather than settled operational prescriptions. They suggest that message receipt should not be treated as sufficient grounds for action in high-consequence operational contexts; that structured verification may be more reliable than unstructured escalation alone; that degraded audio conditions deserve more attention than is often assumed; that exercises should focus on concentrated high-burden scenarios rather than generic awareness alone; and that supplier-related command paths should be incorporated into governance, incident response, and evidence-retention design. These conclusions are compatible with the broader argument of the decision security framework that resilient organisations should not begin with isolated detection tools, but with classification of high-consequence actions, protected verification paths, execution hold rules, and measurable governance arrangements. From an implementation perspective, such safeguards are also more likely to remain sustainable when they are embedded in phased organisational governance models for data-driven systems rather than introduced as isolated technical controls (Jędrasiak & Gawliczek, 2025).

The main scientific contribution of the current study is narrower, but still meaningful. It provides a sector-specific exploratory extension of a broader conceptual framework by showing that, in this power-sector dataset, unsafe outcomes were more clearly associated with behavioural timing than with perceived authenticity and that risk accumulated around a limited set of high-consequence decision scenarios. This empirical extension is important in two ways. Conceptually, it supports the argument that decision security is the correct level of analysis for high-consequence environments. Practically, it identifies a concrete set of power sector contexts in which that

argument matters most. It also places the present work in dialogue with technical deepfake detection literature, which remains indispensable at the artefact assessment level but insufficient at the execution control level when analysed in isolation.

The present study should be interpreted within several important limits. First, only 31 participants generated 744 observations, so the observation count should not be mistaken for an equivalently powered participant sample. Second, repeated observations were nested within participants, which is why clustered modelling was used as a robustness analysis and why condition-level contrasts remain exploratory. Third, the stimulus documentation was designed for scenario, modality, and channel traceability rather than for comparative benchmarking of individual synthesis systems. Fourth, the library scenario was intentionally sector-specific and therefore should not be generalised uncritically to other critical infrastructure domains. These limitations do not invalidate the findings, but they do bound the strength of the conclusions. The current dataset is strongest when interpreted as a behavioural and human-centred operational security study. It is weaker as evidence for broad comparative claims about control effectiveness or as a technical study of generative model realism.

The most defensible conclusion emerging from the present data is that synthetic impersonation in the power sector is dangerous not because recipients necessarily judge synthetic artefacts to be fully authentic, but because plausible high-consequence communications can compress decision time and bypass protected verification logic. The relevant protective task is therefore not limited to better detection. It is the design of operational command pathways in which receipt, verification, authorisation, execution, logging, and supplier escalation are clearly separated and governed. That conclusion is fully consistent with the theoretical position of the earlier decision security framework and provides a strong empirical basis for the governance, implementation, and sector-specific recommendations that follow.

The present findings also help clarify the relationship between detection-oriented deepfake research and organisational security research. Detection-oriented work remains indispensable for understanding media manipulation and benchmark performance (Verdoliva, 2020; Tolosana et al., 2020; Mirsky & Lee, 2021). At the same time, social-impact research shows that deepfakes matter because they reshape trust, uncertainty, and deception in communication environments (Vaccari & Chadwick, 2020; Hancock & Bailenson, 2021). The direct relevance of this literature to cybersecurity has become more explicit in recent work on deepfake-driven social engineering (Pedersen et al., 2025). The present study extends that trajectory into a power-sector context and aligns it with human-centred security research showing that protective behaviour is strongly shaped by organisational design, compliance frictions, and contextual decision demands (Adams & Sasse, 1999; Beautement et al., 2008; Aldawood & Skinner, 2019; Ghafir et al., 2018).

This distinction clarifies the scientific contribution of the study. It allows the analysis to connect deepfake studies, cybersecurity governance, and safety-critical operations without overstating what the current sample can establish. The strongest contribution lies in identifying a behavioural decision-security signal that merits more rigorous, larger-scale follow-up.

Exploratory Implications for Design and Governance

Given the exploratory character of the dataset and the absence of statistically significant between-condition effects, the points below should be read as design implications and hypotheses for future validation rather than as settled operational prescriptions.

Power sector organisations should explicitly map the command paths, approval chains, and coordination routines through which a communication event can trigger technical execution, service disruption, or safety relevant deviation. The present dataset shows that unsafe responses concentrated around selected high-consequence scenarios rather than being evenly distributed across all events. For this reason, controls should be applied most strongly to actions such as work approval, restoration initiation, technically sensitive disclosure, and operational reconfiguration.

A central design principle should be the formal separation of receipt of message from execution of action. No high-consequence operational instruction should be treated as executable solely because it was received through a familiar channel or appeared to originate from an authoritative sender. Callback verification, threshold authorisation, and execution hold logic should be defined in such a way that communication cannot be transformed directly into action without protected routing.

The results indicate that verification was associated with the lowest unsafe response burden among the four main behavioural categories. This suggests that organisations should strengthen structured verification pathways rather than relying on generic escalation alone. Escalation remains necessary, but it should be directed to the correct actor, through the correct channel, and in a way that preserves control over whether the requested action may proceed.

Supplier related and contractor mediated communication should be governed as part of the operational decision security architecture. Where external actors can request access, initiate coordination, influence timing, or relay technically consequential instructions, organisations should require named escalation contacts, documented callback procedures, logging obligations, notification clauses, and evidence retention expectations. Supplier assurance should therefore be treated as an operational safeguard rather than a secondary procurement formality.

The elevated unsafe response burden observed in noisy audio conditions, together with the significantly shorter decision times associated with unsafe responses, indicates that exercising only idealised communication conditions is insufficient. Training and preparedness activities should include degraded signal quality, time pressure, and authority loaded scenarios. At the same time, organisations should measure decision time, unsafe execution rate, and verification initiation rate as practical indicators of whether decision security is improving under realistic operational pressure.

Limitations and Future Research

The present study should be interpreted within several limitations. First, the design was scenario-based and quasi-experimental rather than incident based, which means that the reported unsafe response burden should not be treated as a direct estimate of real world event frequency. Second, although condition effects were directionally favourable, several global comparisons did not reach conventional thresholds of statistical significance. Accordingly, the present dataset supports a careful operational interpretation, but not a definitive claim that one control configuration is universally superior under all power sector conditions. Third, the participant sample, while adequate for an exploratory sector-specific study, remains moderate in size and should be expanded in future work across a broader range of operators, transmission entities, distribution actors, and supplier facing teams. Future research should extend the dataset in four directions. The first is larger-scale replication with additional organisations and a wider role spectrum. The second is preregistered comparison of control regimes with participant-level power targets. The third is more detailed publication of stimulus documentation, including modality metadata, channel conditions, scenario linkage, and independent realism checks. The fourth is deeper modelling of supplier-mediated command chains and their interaction with time-pressured operational decision making.

Conclusion

This exploratory study does not establish statistically conclusive superiority of any single procedural protection regime. Its strongest empirical contribution is different: unsafe responses were associated with shorter decision times, perceived authenticity did not separate safe from unsafe outcomes, and unsafe responses concentrated in a limited set of high-consequence power-sector scenarios. These patterns are more consistent with a decision-security interpretation than with a purely media-authenticity interpretation. The dataset showed that more than one quarter of tested decision events resulted in an unsafe response at the observation level. This burden was not

distributed uniformly. It accumulated around technically sensitive requests involving disclosure, work approval, restoration logic, and operational reconfiguration. That concentration is important because it suggests that the most relevant unit of protection is not the abstract category of synthetic media, but specific decision points at which communication intersects with authority, urgency, and technical consequence. The study also found a descriptively favourable gradient across procedural conditions, with lower unsafe response burden under stronger control configurations. However, those between-conditioning were not statistically significant in the present dataset and should therefore be treated as hypothesis-generating rather than conclusive. A further important finding is that unsafe outcomes were not meaningfully distinguished by authenticity judgements. By contrast, unsafe decisions were associated with faster response times, and that signal remained significant in a participant-clustered model. This suggests that the more defensible explanation for unsafe responding in the present study is behavioural compression at the transition from message receipt to action selection. For power-sector governance and engineering, the present article therefore offers a cautious agenda rather than a final prescription. It supports closer examination of high-consequence command paths, stronger separation of receipt from execution, more explicit governance of supplier-mediated communication, and exercise designs that account for degraded channels and decision speed. At the scientific level, the study contributes by reframing synthetic impersonation in the power sector as a human-centred operational security problem and by identifying a concrete agenda for future work on command integrity, supplier-mediated trust chains, and time-pressured decision behaviour. Larger and more technically transparent replications are now required to determine whether the exploratory design implications observed here can be converted into robust operational evidence.

Funding

This publication was prepared as part of the project “Development of an Interpretable Deepfake Detection Algorithm Using a Deepfake Feature Atlas”, no. 145/2025/2, co-funded by the Metropolitan Fund for Supporting Science Programme of the Górnośląsko-Zagłębiowska Metropolis.

Competing interests

The authors declare that they have no competing interests.

Use of Artificial Intelligence

Artificial intelligence tools were used solely for technical text checking.

References

Adams, A., & Sasse, M. A. (1999). Users are not the enemy. Communications of the ACM, 42(12), 40-46. https://doi.org/10.1145/322796.322806

Aldawood, H., & Skinner, G. (2019). Reviewing cyber security social engineering training and awareness programs: Pitfalls and ongoing issues. Future Internet, 11(3), 73. https://doi.org/10.3390/fi11030073

Beautement, A., Sasse, M. A., & Wonham, M. (2008). The compliance budget: Managing security behaviour in organisations. In Proceedings of the 2008 New Security Paradigms Workshop (pp. 47-58). https://doi.org/10.1145/1595676.1595684

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., Dafoe, A., Scharre, P., Zeitzoff, T., Filar, B., Anderson, H., Roff, H., Allen, G. C., Steinhardt, J., Flynn, C., Ó hÉigeartaigh, S., Beard, S. J., Belfield, H., Farquhar, S., Lyle, C., Crootof, R., Evans, O., Page, M., Bryson, J., Yampolskiy, R., & Amodei, D. (2018). The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. arXiv preprint arXiv:1802.07228.

Buchwald, P., & Jędrasiak, K. (2024). A method to protect users from fake news disinformation content using hybrid evaluation methods based on artificial intelligence solutions and user experience. Safety & Fire Technology.

Cherdantseva, Y., Burnap, P., Blyth, A., Eden, P., Jones, K., Soulsby, H., & Stoddart, K. (2016). A review of cyber security risk assessment methods for SCADA systems. Computers & Security, 56, 1-27. https://doi.org/10.1016/j.cose.2015.09.009

Chesney, R., & Citron, D. (2019). Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. California Law Review, 107, 1753-1819.

Cichonski, P., Millar, T., Grance, T., & Scarfone, K. (2012). Computer Security Incident Handling Guide (NIST Special Publication 800-61 Rev. 2). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-61r2

Egelman, S., & Peer, E. (2015). Scaling the security wall: Developing a security behavior intentions scale (SeBIS). In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (pp. 2873-2882). https://doi.org/10.1145/2702123.2702249

Ghafir, I., Saleem, J., Hammoudeh, M., Faour, H., Přenosil, V., Jaf, S., Jabbar, S., & Baker, T. (2018). Security threats to critical infrastructure: The human factor. The Journal of Supercomputing, 74, 4986-5002. https://doi.org/10.1007/s11227-018-2337-2

Hancock, J. T., & Bailenson, J. N. (2021). The social impact of deepfakes. Cyberpsychology, Behavior, and Social Networking, 24(3), 149-152. https://doi.org/10.1089/cyber.2021.29208.jth

Jędrasiak, K. (2024a). Analiza cyberzagrożeń współczesnych kanałów komunikacyjnych ze szczególnym uwzględnieniem zagrożeń typu DeepFake. In Horyzonty sztucznej inteligencji a Przemysł 5.0 (pp. 37-59). Wydawnictwo Naukowe Akademii WSB.

Jędrasiak, K. (2024b). Audio stream analysis for deep fake threat identification. Civitas et Lex, 1(41). https://doi.org/10.31648/cetl.9393

Jędrasiak, K. (2025). Analiza cech jakości obrazu oraz artefaktów przetwarzania jako fundament detekcji deepfake. Safety & Fire Technology.

Jędrasiak, K. (2026). Decision security against synthetic impersonation in high trust sectors. European Cybersecurity Journal, 11(1), 53–66.

Jędrasiak, K., & Bijoch, J. (2025). Deep network representations as reliable indicators of synthetic content in audiovisual and clinical contexts. Communications of International Proceedings, IBIMA 46.

Jędrasiak, K., & Gawliczek, P. (2025). Implementing artificial intelligence in data-driven enterprises through a ten-phase framework based on multiple case studies. Social Development and Security, 15(5), 17-34.

Jędrasiak, K., & Wolański, R. (2023). Audio-video analysis method of public speaking videos to detect deepfake threat. Safety & Fire Technology, 62(2), 172-180.

Jędrasiak, K., & Wolański, R. (2024). An image analysis algorithm for detecting modified content on the internet against cybercrime: A technique for estimating the probability of modification. Safety & Fire Technology, 63(1), 88-94.

Kietzmann, J., Lee, L. W., McCarthy, I. P., & Kietzmann, T. C. (2020). Deepfakes: Trick or treat? Business Horizons, 63(2), 135-146. https://doi.org/10.1016/j.bushor.2019.11.006

Knowles, W., Prince, D., Hutchison, D., Disso, J. F. P., & Jones, K. (2015). A survey of cyber security management in industrial control systems. International Journal of Critical Infrastructure Protection, 9, 52-80. https://doi.org/10.1016/j.ijcip.2015.02.002

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

Marshall, N., Sturman, D., & Auton, J. C. (2024). Exploring the evidence for email phishing training: A scoping review. Computers & Security, 139, 103695. https://doi.org/10.1016/j.cose.2023.103695

Mirsky, Y., & Lee, W. (2021). The creation and detection of deepfakes. ACM Computing Surveys, 54(1), 1-41. https://doi.org/10.1145/3425780

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), Article 18. https://doi.org/10.3390/jcp5020018

Rose, S., Borchert, O., Mitchell, S., & Connelly, S. (2020). Zero Trust Architecture (NIST Special Publication 800-207). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-207

Stouffer, K., Pease, M., Tang, C. Y., Zimmerman, T., Pillitteri, V., Lightman, S., Hahn, A., Saravia, S., Sherule, A., & Thompson, M. (2023). Guide to Operational Technology (OT) Security (NIST Special Publication 800-82 Rev. 3). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-82r3

Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., & Ortega-Garcia, J. (2020). Deepfakes and beyond: A survey of face manipulation and fake detection. Information Fusion, 64, 131-148. https://doi.org/10.1016/j.inffus.2020.06.014

Vaccari, C., & Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. Social Media + Society, 6(1), 1-13. https://doi.org/10.1177/2056305120903408

Verdoliva, L. (2020). Media forensics and deepfakes. IEEE Journal of Selected Topics in Signal Processing, 14(5), 910-932. https://doi.org/10.1109/JSTSP.2020.3002101

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

Workman, M. (2008). Wisecrackers: A theory-grounded investigation of phishing and pretext social engineering threats to information security. Journal of the American Society for Information Science and Technology, 59(4), 662-674. https://doi.org/10.1002/asi.20779