

**JOURNAL OF INFORMATION
SYSTEM AND TECHNOLOGY
MANAGEMENT
(JISTM)**www.gaexcellence.com/jistm**NAVIGATING CYBERSECURITY AND TRUST BARRIERS:
A QUALITATIVE STUDY OF AI MANAGEMENT IN THE
UAE SECURITY SECTOR**Khalfan Abdulla^{1*}, Rafiduraida Adulrahman²¹Faculty of Management and Economic, Sultan Idris Education University, Malaysia K.uae.dxb@gmail.com <https://orcid.org/0009-0003-2897-7817>²Faculty of Management and Economic, Sultan Idris Education University, Malaysia rafiduraida@fpe.upsi.edu.my <https://orcid.org/0000-0001-8225-188X>

*Corresponding Author

Article Info:**Article history:**

Received date: 19.07.2026

Revised date: 18.08.2026

Accepted date: 10.09.2026

Published date: 28.09.2026

To cite this document:

Abdulla, K., & Adulrahman, R. (2026). Navigating Cybersecurity And Trust Barriers: A Qualitative Study Of Ai Management In The UAE Security Sector. *Journal of Information System and Technology Management*, 11 (44), 621-639.

Abstract:

This qualitative research paper explores the interaction between cybersecurity vulnerabilities and trust barriers influencing artificial intelligence (AI) management practice in the United Arab Emirates (UAE) security industry, addressing a gap left by quantitative acceptance studies that have treated these barriers as independent variables rather than as interacting factors. The main conclusion is that low trust and cybersecurity vulnerability are not independent hindrances but support each other. AAn interpretivist, exploratory design used a purposive sampling strategy and semi-structured interviews with 35 participants comprising security professionals, AI managers, cybersecurity experts, and governance staff, and data were analysed thematically. The results indicate that AI-related vulnerabilities, particularly adversarial manipulation and data-integrity breach create opaque, unexplainable AI behaviour, to which individuals adjust with a contingent, multidimensional type of doubt instead of a broad-based technological scepticism. In security, governance has not yet been formalised as fast as AI deployment in comparison to more mature governance in service delivery fields like healthcare and municipal services and the most viable management solutions participants identify are hardening security and accountability-making instead of two distinct initiatives. The study introduces a reconceptualisation of trust in high-stakes AI environments that is founded on multidimensional, offering technological, human, and organisational fields and recommends that governance formalisation should be seen as a step toward, rather than a follow-up to AI deployment by UAE security organisations and regulators.

DOI: 10.35631/JISTM.1144038

Keyword:

AI Governance, Artificial Intelligence, Cybersecurity, Qualitative Research, Security Management, Trust, United Arab Emirates



© The authors (2026). This is an Open Access article distributed under the terms of the Creative Commons Attribution (CC BY NC) (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact jistm@gaexcellence.com.

Introduction

The international security field has plunged into a moment where artificial intelligence (AI) is ceasing to be merely an adjunct tool, but constitutive element of operational decision-making and this transition has implications far beyond the technical performance metrics that have been predominant in the literature so far. Surveillance analytics, predictive threat modelling, biometric authentication, and automated incident response is now the focus of where security organisations devote efforts and resources, but it is the very aspect that provides AI with an appeal to security work, which is its ability to act upon patterns far beyond human perceptual bandwidth, that makes its outputs challenging to audit, challenge, and trust (Al Jaberi et al., 2023; Awad et al., 2024). This tension is not accidental to the adoption of AI; it is structural, as it is the same opaque that gives operational advantage that is the one that takes the conditions away under which the human overseers can check whether the advantage is being deployed in ways that are not inappropriate. In turn, the tendency to conceive cybersecurity risk and trust deficit as two parallel, yet separate issues, as most quantitative adoption research implicitly does, clouds the more crucial empirical question of how the two are interrelated in a particular management practice.

In the United Arab Emirates, this dynamic is intensified by a policy environment that has turned the use of AI into an explicit tool of state development as opposed to a discretionary efficiency enhancement. The deployment of AI in national strategy through the strategies of the UAE, which appear to be sector-cutting, including healthcare, government communication, oil and gas, and municipal services, has resulted in a relatively rapid proliferation of AI tools into publicly visible and security-related tasks (Al Mazmi & Radwan, 2026; Al Sulaity et al., 2025; Alyileili & Opoku, 2026). However, top-down strategic mandate-driven rapid diffusion is qualitatively different than bottom-up organisational demand-driven diffusion, since the former pushes the time when institutions gain the governance, auditing, and workforce capabilities that will render AI trustworthy in practice (Alshehhi et al., 2024). This compression especially affects security organisations: where a municipal self-service portal has a substantial effect that can be reversed, the impact of an AI system failing or being compromised cannot be undone, and so the governance catch-up margin in this sector is smaller than in most other areas where the adoption of AI in the UAE has been examined.

The critical analysis of the available literature demonstrates that cybersecurity and trust have up to this point been analyzed more as independent research programmes than as a pair of interacting variables in the context of a single management issue. The literature on

cybersecurity has paid close attention to technical vulnerability: manipulation of the detection and tracking systems by adversaries through the means of AI (Al Jabeti et al., 2023). This exposes the SCADA-connected critical infrastructure to AI-based attack and defence dynamics (Alattas Alhashmi et al., 2023), and the legal exposure that is compounded in the legal sphere when AI and IoT are implemented in the regulated sectors, like the financial one (Alazzam et al., 2026).

In contrast, the trust literature has focused on perception: attitudes of the population towards AI-enabled drone delivery (Almuraqab et al., 2025), the acceptability of autonomous vehicles (Hafiz et al., 2025), and procedural fairness that facilitates confidence in policing (Sebire et al., 2025). What is missing to a great extent is an explanation of how the exposure to cybersecurity and the development of trust mutually constitute each other in the day-to-day management practice of security workers who have to protect the systems and simultaneously defend their usage to the disbelieving co-workers, the supervisory bodies, and the general population. This is the gap the current research targets, and it is a gap of theoretical synthesis and not of empirical novelty per se, as much of the underlying phenomena have been reported singly but seldom in a single framework of analysis.

The methodological imbalance in the UAE AI literature that is also worth mentioning, rather than as a minor limitation of previous literature, is also a motivation for this study. Structural equation modelling and other quantitative methods to measure the acceptances based on the technology-acceptance theory have been the dominant analytic approaches in UAE-centered studies on the adoption of AI (Alhashmi et al., 2020; Al-Skaf et al., 2021). These methods are quite appropriate to testing the hypothesis that perceived usefulness predicts adoption intention but are inappropriate to test the cause of a security professional with a high score on perceived usefulness not operational trusting a system whose decision making they cannot interrogate. Quantitative acceptance models assume that trust is a scalar attitude, which may vary in intensity; qualitative explanations of the role of trust in professional practice assume, instead, that trust in high-stakes situations of security is better thought of as a conditional, situationally renegotiated judgment, contingent on accountability frameworks as much as on system performance perceptions (Alrasheed et al., 2025). The current research thus takes a qualitative design not as a methodological choice but as a methodologically required one, considering that the research questions deal with meaning-making and situated judgment and not the magnitude of association between preconceived constructs.

In line with this, the purpose of the research is to understand how the impact of cybersecurity vulnerabilities and trust barriers can be combined to influence AI management practice in the UAE security industry, specifically focusing on those points where the technical risk and relational confidence meet in the experiences of practitioners. The result of this objective is four objectives, first, to determine the cybersecurity problems that security professionals correlate with the implementation of AI, second, to explore the barriers to trust that precondition the implementation and further use of AI, third, to comprehend how stakeholders perceive the use and governance of AI, and fourth, to describe the strategies practitioners report to be implementing to address both risk and scepticism at the same time.

These goals are structured around one overarching research question, which is how cybersecurity and trust barriers affect AI management in the UAE security industry, and three minor questions, including the cybersecurity risks that AI implementation entails, the trust perceptions of the stakeholders and the management strategies to address the two. The

importance of this question is three-fold: it provides a qualitative, practice-level explanation to a body of literature that is still largely quantitative in its acceptance modelling. It also provides security organisations with an empirically informed set of guidelines on how to design governance mechanisms in response to trust and security as interrelated, rather than sequential, issues. Lastly, it targets UAE policymakers who are formalising AI governance frameworks today with a body of evidence in security contexts that is still sparse compared to evidence in healthcare or municipal service contexts (Alyileili & Opoku, 2025a; Khalil & al-Ali, 2026). The rest of the article follows in the following manner: Section 2 outlines the methodology of the qualitative research and its rationale; Section 3 summarizes the thematic results; Section 4 compares the thematic results with the rest of the literature and theorizes on the implications; and finally, Section 5 concludes with recommendations on how the governance practice should be.

Literature Review

Methods

Research Design and Philosophy

The design used in this study is an exploratory qualitative study design, and the decision to use this type of study design is based on the above conceptual gap and not as a result of a general inclination towards qualitative research studies in preference to quantitative research studies. The exploratory designs are suitable in the case when the variables are not well defined and the relationship between them are not defined in the literature yet, and this is precisely the situation that takes place here: although all the cybersecurity risk factors and trust antecedents have been listed down separately in previous UAE-specific studies, the nature of their interaction with each other within the context of security-sector AI management has not been theorised to a sufficient degree to facilitate hypothesis-testing research (Al Jaber et al., 2023; Alrasheed et al., 2025). The design is based on an interpretivist approach to research, which assumes that cybersecurity risk and trust is not an objective feature of AI systems that exists independently of the individuals who should manage it, but rather constituted by the meaning given to system behaviour, institutional accountability, and professional responsibility by security professionals. A direct methodological implication of this philosophical position is that data collection methods that can reach situated interpretation have to be used, not just behavioural or attitudinal scores, which explains the choice of semi-structured interviewing, as opposed to survey instrumentation, as the main data-generating technique.

Study Setting

The research is framed within the context of security sector within the UAE in a broader sense, which includes both state security services, i.e., police and civil defence agencies that have implemented AI-based surveillance and threat-detection systems (Alattas Alhashmi et al., 2023; Sebire et al., 2025), and private security and critical-infrastructure organisations that have adopted AI in operational risk management. This sector is estimated to have about 3500 personnel. This two-locus model was not imposed on a single type of organisation but was adopted to avoid confounding results that are causes and effects of the culture of government bureaucracy with results that can be attributed to the security role itself, and vice versa: a wholly public-sector sample would lose the unique accountability pressures that characterise the state security work. The UAE situation is analytically relevant beyond the convenience of

access: the centralised form of governance and high state capacity to invest in technology creates conditions of AI diffusion that are distinctly different than the more contentious (Abulbasher, 2026). Multi-veto-point processes of AI diffusion that are recorded in liberal-democratic security contexts, implying that results produced here are relevant to the governance configuration that is today under-represented in the international AI-security literature.

Sample Size Determination

Slovin formula: $n = N / (1 + N(e^2))$

Estimated population (N)=3,500 AI/security personnel in UAE organizations; margin of error=15%; resulting sample size=35 respondents.

Participant Selection

The participants were identified by using purposive sampling, and they were selected as those individuals who would speak with first-hand authority on the intersection of cybersecurity and trust in AI management, and not on AI adoption, in general. Inclusion criteria were a set of four overlapping professional categories: security professionals who have direct operational experience of AI-enabled systems; AI managers who have to deploy and manage AI systems in security organisations; cybersecurity experts who have to identify and address AI-specific vulnerabilities; and policy or governance officials who have to translate national AI strategy into sector-level regulations. This stratified purposive design was needed in that each category is better placed to respond to a different aspect of the cybersecurity-trust relationship: the operational staff members are best able to explain how trust is implemented or not in their everyday work, the cybersecurity experts to explain the technical vulnerability environment, and the governance members to explain the institutional mechanisms that are supposed to balance the two. The sampling was continued until thematic saturation was perceived to have been attained in these categories in accordance with the qualitative practice of thematic exploration that is known to exist concerning the use of an exploratory approach to thematic analysis.

Data Collection

Semi-structured interviews were used to collect data, a technique that is appropriate due to its ability to offer comparability among participants with enough freedom to pursue unforeseen lines of disclosure, which is relevant in a field where participants may be bound by organisational norms on confidentiality on security operations. The interview guide was structured into four thematic anchors, which include the direct experience of the participants with the implementation of the AI within their organisation; the cybersecurity concerns that they related to the implementation; the perceptions of trust and mistrust that they held and their colleagues held about the outputs and recommendations of AI; and the management practices that their organisation had been undertaking, either formally or informally, to deal with both the concerns. Probing questions were employed to shift participants off of general statements to more specific, situated examples, based on the assumption that abstract attitudinal statements are not as helpful analytically as descriptions of specific decisions, failures, or instances of doubt.

Data Analysis

The thematic analysis was used to analyse the data, following the by now standard process of familiarisation, initial coding, theme generation, theme refinement, and interpretation. The analysis was also clearly abductive instead of purely inductive: although the codes were generated based on the data instead of being imposed upon them by an existing framework, the refinement and interpretation phase was carried out through active discussion with the reviewed literature on cybersecurity and trust, allowing emergent codes to be read against, and in some cases challenge, existing theoretical anticipations. This abductive position is critical to make quite clear since a mere inductive assertion of theory-free coding is hardly practical in the field, and disclosure of interpretive frame empowers instead of undermines the plausibility of the emerging themes.

Quality and Trustworthiness of Research

Reliability was dealt with by the four standards, which are traditionally used in the qualitative study credibility, transferability, dependability, and confirmability. The stratified purposive sampling strategy provided credibility because it enabled cross-checking of the accounts across the various professional categories as opposed to using one occupational perspective. Transferability is necessarily limited: results are reported as analytically generalisable to security organisations in similar circumstances of high-investment, centralised-governance, as opposed to statistically generalisable to the UAE security sector as a population. Dependability was sought by writing down the decision to make the coding, and confirmability by being reflexive about the role of the researcher in the creation of the themes, especially with the above abductive analytic position.

Analysis and Findings

Across the four themes presented below, cybersecurity vulnerability and low trust in AI systems emerge not as independent problems but as mutually reinforcing ones: unresolved cybersecurity risk deepens practitioner distrust, and that distrust in turn shapes how security organisations respond to risk. This association, which will also be found in full at the end of this part, is the key discovery of the work and is presented here prior to discussing the sample of participants and the four themes.

Participant Profile

In line with the stratified purposive sampling approach, the sample of participants among the four targeted professional types was distributed unevenly, with slightly higher representation of operational and cybersecurity professions, due to their relatively higher exposure to AI systems in their daily activities than policy personnel, who were still included due to their unique perspective on the design of governance. The tenure and seniority of the participants in the organisation were diverse, which was analytically helpful to the extent that the participants who were older in the organisation were more inclined to frame the issue of trust in institutional and reputational terms, whereas the participants who were on the front-line tended to frame the issue of trust in operational and procedural terms, which is also a finding in its own right, not just the demographic background.

Table 1: Participant Profile

Category	Frequency	Percentage (%)
Security Professionals	12	34.3
AI Managers	8	22.9
Cybersecurity Specialists	10	28.6
Governance Personnel	5	14.3

Table 2: Professional Experience Distribution

Experience	Frequency	Percentage (%)
1–5 years	9	25.7
6–10 years	13	37.1
11–15 years	8	22.9
16+ years	5	14.3

Theme 1: AI Systems and The Emergence of Cybersecurity Risks

The initial theme that appeared in the data results is that the participants did not sense AI-related cybersecurity risk as a homogenous threat but as a stratified set of different types of exposure, all of which have different implications on management response. The sub theme most frequently mentioned was weaknesses inherent in AI infrastructure per se, especially detection and tracking systems being vulnerable to adversarial manipulation. Respondents in cybersecurity roles cited not just fear of outside intrusion in the traditional sense, but a more insidious threat. They cautioned that the AI system will be manipulable to miss detecting the inputs, with no classical violation of the system itself. Technical literature confirms this threat on generative adversarial threats with object tracking and surveillance systems (Al Jaber et al., 2023). It can be used to reduce the quality of detection by adversarial perturbation, without causing concrete harm to the underlying infrastructure. A story given by one participant of an internal red-teaming exercise where they had successfully subverted the accuracy of a detection system without any conventional alert to provide would demonstrate a stronger interpretive argument: that the threat model that security organisations have traditionally been based on (which is based on unauthorised access and data exfiltration) is ill suited to the threat model where the inferential logic of the system turns into the attack surface. This discrepancy between old-model threats and AI-specific vulnerabilities profiles repeated in the accounts of the participants and is a more fundamental result than an individual vulnerability profile.

As one participant illustrated:

“In a recent red-team exercise, the attackers did not compromise our firewalls or steal data, but instead, they just injected adversarial noise into the camera feeds. The AI did not even recognize that we were under attack at all, and our traditional threat framework was entirely unaware that we were being attacked.” (Cybersecurity Specialist, 8 years experience).

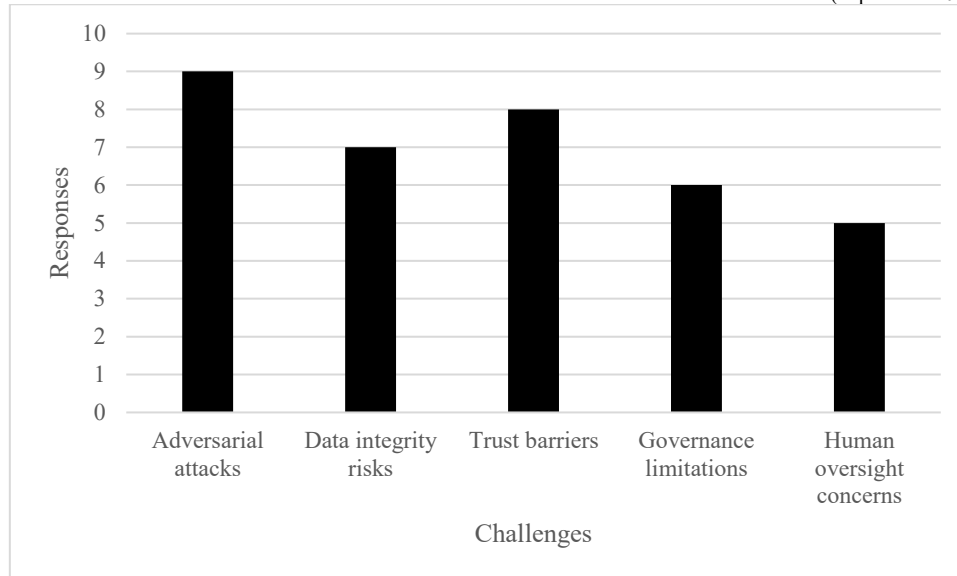


Figure 1: Interview Responses from UAE Security-Sector Participants

Theme 2: Ai Trust Issues

The second significant theme is trust, and the evidence indicates that lack of trust in this respect is not in the first place a product of the lack of familiarity with the AI-technology, which is the implicit premise of much of the technology-acceptance research, but rather a product of a lack of transparency in a specific sense: an inability to obtain justification for a system's outputs after the fact. Respondents consistently made the difference between not knowing how a system operated, which they were usually comfortable with, and not knowing how to obtain a justification of a given output, when necessary, which they were not comfortable with. This differentiation redefines the long-established fact that a lack of transparency erodes AI confidence (Ambali Parambil et al., 2024) in a more specific sense: transparency in the abstract meaning of explainable model architecture does not seem like a key concern to practitioners as compared to transparency in the operational sense of after-the-fact accountability, that is, the possibility to reconstruct why a particular alert or recommendation was produced when its impact is under examination by a supervisor, a regulatory agency, or in the most severe cases a court or an arbitration panel.

One operational leader argued that

“This does not mean that I will need to know the complicated math behind the neural network, but in case my team triggers a critical response on the basis of an AI flag, and a regulatory agency later demands that I explain why I took that specific action, I will have no opportunity to answer with the phrase that the machine told me to do that.” (Security Professional, 14 years experience)

The result of this finding is in line with larger body of public-attitude research on AI-enabled security technologies in UAE context that has determined that people are not unconditionally accepting of AI-enabled security applications, but this varies depending on the perceived stakes and reversibility of the decision under consideration (Almuraqab et al., 2025; Hafiz et al., 2025). What the current data contribute to this public-facing result is a practitioner-side mirror image of the same conditionality: practitioners seem to use a similar stakes-sensitive calibration

to their own readiness to be misled by AI outputs, indicating that the conditional-trust pattern observed among the general public is not merely a lay misperception that can be resolved through campaigns to educate the public about AI.

Theme 3: Governance and Management Issues

The third theme is the governance frameworks in which security organisations seek to operate concurrently to control cybersecurity and trust and the main finding here is that actors felt that there was a consistent gap between the speed of AI capability implementation and the speed of formalisation of governance. The most frequently mentioned subtheme was policy restrictions. Regulatory guidance on sector specific AI applications to security uses was reported to be less developed by the governance personnel, including than regulatory attention in areas already facing the public like healthcare and municipal services. The sectors were selected to be compared since they have longer-standing deeper public transparency requirements that have historically led to quicker formalisation of governance. This contrast educates on what hastens or slows governance maturity in other places. The stories shared by the participants agree with the existing body of literature on the topic of sector governance (Al Badi et al., 2022; Alyileili and Opoku, 2025a). Such stories show that the level of the UAE AI governance distribution does not evenly cover sectors. It is on the contrary concentrated where there is a focus on public facing service delivery on strong incentives of visible governance. This exposes security applications to being relatively ungoverned whereas there is a relatively lack of risk profile that would justify similarly or more governance attention.

The second subtheme, regulatory uncertainty, was characterized by participants not so much as a lack of regulation, but as a lack of settled interpretation as to how the existing rules applied to truly new AI-specific situations, a result that is consistent with the legal-scholarly literature that records that the doctrines of liability, such as strict liability, negligence, and vicarious liability, all do not calibrate well to harms caused by autonomous decision systems (Ibrahim Abdulrahim Abdulla et al., 2025). A number of respondents spoke of internal practice where judgments on how to address AI-related incidents were made in an informal and unpredictable manner across cases exactly due to the lack of any settled doctrine to guide them, which is a more disturbing result than mere regulatory gaps, as inconsistent informal practice can become a source of organisational risk and inequity on its own before any formal legal exposure comes into being.

A third subtheme was human oversight requirements, with participants describing human oversight here in interesting ways compared to the formal language of governance used in international AI ethics frameworks, which generally define human oversight as a fixed structural factor, such as the existence of a human reviewer at a decision point (Mohammed et al., 2025). Instead, participants reported oversight being calibrated to the nature and intensity of each AI function, and that standardised human-in-the-loop requirements, though administratively straightforward to articulate, might not fit the heterogeneous risk profile of AI functions in a single security organisation, and in certain instances may result in a culture of check-box compliance that meets the formal requirement but not the substantive oversight it is intended to achieve.

The fourth subtheme, organisational readiness, reflected the perception of the participants that the process of formalisation of governance was accelerating faster than the organisational and cultural readiness to realise it, which is also in line with larger UAE-public-sector evidence that

the success of the procurement and adoption rely heavily on the organisational and human-capital readiness factors that are often underweighted compared to the technical and procedural ones (Alshehhi et al., 2024). A number of the participants observed that the new internal AI governance policies were on paper, without a significant investment in the training to make them meaningful, a trend that relates this theme directly to the strategies described in Theme 4.

One government official observed the practical drawbacks of blanket policies:

"We have policies to have a human in the loop when deploying AI, but in practice, the AI processes threat data so fast that the human reviewer is simply clicking approve without the time or ability to actually review the data. It gives people a false sense of security, which is check-box compliance, not substantive oversight" (Governance Personnel, 11 years' experience).

Theme 4: AI Secure and Trust Building Strategies

The fourth theme describes the practices participants reported, officially authorized or unofficially ad-hoc, to deal with the cybersecurity and trust issues identified above, and the most notable trend in this theme is that those strategies that appeared to be most acceptable in the eyes of participants were those that dealt with cybersecurity and trust together and not in parallel but separate efforts. Human-centred AI management became the most popular subtheme, and participants noted a conscious attempt to retain a human decision-owner on any action that has a material impact and is influenced by AI, which participants explicitly explained by the accountability gap that has been identified in Theme 3 instead of making it a generic best-practice gesture. This result, not merely confirms but extends, the wider evidence base of the UAE that human-centred design positively influences the legitimacy of technology-enabled interventions to the public (El Naggar & Abdelrazik, 2024). This is through demonstrating that in the security domain in particular, the benefit of the legitimacy of human-centred design does so significantly via its role as a liability-bearing accountability anchor and not just via its impact on user experience or perceived control.

The second subtheme was cybersecurity frameworks, where participants reported trying to apply zero-trust architectural principles, traditionally designed to manage conventional network security, to the interactions between AI elements and the overall security infrastructure, a trend that aligns with recent technical research indicating the generalizability of zero-trust models to distributed and ad-hoc network environments (Barbosa et al., 2025). The adoption of this framework by participants is analytically important in that it is a manifestation of organisations understanding that AI components are to be treated as untrusted components of the network and continuously checked, rather than as trusted extensions of existing trusted infrastructure. A conceptual shift that is antithetical to the more common organisational tendency of treating a newly purchased AI system as an upgrade to existing trusted infrastructure as opposed to a new attack surface requiring independent checking.

The third subtheme, staff training and awareness, has been defined by the participants in a non-generic way, as digital-literacy training, with a focus on scenario-based training, in which the participants should have been trained on how to identify AI-specific failure modes, such as adversarial manipulation and edge-case misclassification, instead of being simply taught how to use the AI tool itself. This difference is significant as the current literature on the topic of

workforce-readiness in the UAE mostly models AI training as a skills-training or change-management issues (Khalil & al-Ali, 2026; Shwede, 2024). The respondents in security roles were specifically describing training as a way to sustain a healthy professional scepticism, a training outcome that is in a way of conflict with the usual change-management literature, which focuses on developing confidence and minimizing resistance as its major training goal, and this conflict is itself worth highlighting as an unsolved design issue in the design of AI-training programme in the security sector.

Respondents were particularly explaining training as a means of maintaining a sound professional scepticism. One AI Manager remarked that

“Our training is not on learning to use the new dashboard, but on scenario-based exercises to demonstrate to the operators how exactly the AI can become hallucinated or manipulated through edge cases. We are letting them use the tool, but we are actively training them not to trust it blindly” (AI Manager, 7 years experience)

The fourth subtheme, accountability mechanisms, was defined as the structural complement to the above human-centred management practices and with several participants mentioning internal audit trails, compulsory incident review processes, and (most institutionally developed) internal discussion of how arbitration-style accountability mechanisms suggested in the legal literature on AI arbitrator liability (Ibrahim Abdulrahim Abdulla et al., 2025) could be redesigned to suit internal security-incident review.

Summary of Findings

When combined, the four themes intersect in a single overarching pattern which is greater than the sum of its parts: cybersecurity vulnerability and lack of trust in UAE security-sector AI management is not a parallel issue in need of parallel solutions, but rather one that reinforces the other, so that the failure to tackle cybersecurity risk, especially risk created by opaque, difficult-to-explain system behaviour, directly creates the accountability gap practitioners experience in the form of a trust deficit, and the lack of trust that creates the lack of governance vulnerability respectively. This circular relation, neither of the factors alone, is the key empirical result of this study, and the analytic basis of the discussion, which follows.

Discussion

Interpreting Cybersecurity Challenges

The cybersecurity results presented in Section 3 are consistent with, yet significantly expand, the wider global discussion on AI-specific security risk. Adversarial manipulation, data-integrity breach, and integration vulnerability are the three major categories of AI-specific risk that are not similar to conventional cybersecurity risk, which have been established with a reasonable consensus in the international literature (Al Jaber et al., 2023; Awad et al., 2024). The contribution of the current results is support that, in the UAE security industry in particular, the weightings of these risk categories are not balanced in terms of focus by the practitioners; integrity-related issues were more prevalent than confidentiality-related issues in the participants description, which is contrary to the relative emphasis in general-purpose AI-adoption literature, where confidentiality and data-privacy concerns are generally prioritized (Bahgat, 2025). The divergence is not best described as an inconsistency but rather as a

consequence of sectoral risk profile: in a security context, the operational impact of a silently degraded inference is plausibly worse than the impact of disclosing data, which would justify why the lived risk hierarchy of practitioners is not equal to the hierarchy implicit in general AI-governance discourse. This is a conceptually useful result in that it is proposing the idea that risk-prioritisation models that are formulated in general-purpose AI governance cannot simply be applied to the security-sector domains without re-weighting to sector-specific consequence profiles.

Understanding Trust as a Determinant of AI Management

The trust results serve to reconceive a conceptualisation of trust that leaves behind the hegemonic approach to treatment of trust as a one-dimensional attitudinal measure in technology-acceptance studies. The current evidence rather confirm a multidimensional explanation where trust acts at the same time at human level, professional judgment on when to wait to a system; at the technological level, trust on a system to be interrogated; and at the organisational level, trust on the existence of accountability structures in institutions to settle the conflict in case of challenges in system outputs. This multidimensional framing directly theoretically buys such an apparent paradox, as is evident in a number of UAE adoption studies, where perceived usefulness scores are high whilst actual operational trust scores are conditional or partial (Alhashmi et al., 2020; Almuraqab et al., 2025). When trust is operationally multidimensional, in the way that the current data indicate, then the apparently high scores on a usefulness-oriented acceptance scale, which mainly reflects the technological dimension, should be accompanied by low scores on the organisational-accountability dimension, as the two dimensions are conceptually and empirically distinct and not two measures of the same underlying attitude. This reframing implies that quantitative acceptance studies in this field in future would be enhanced by explicitly separating the dimensions of trust into these component dimensions instead of considering it as a single latent construct because collapsing the dimensions may obscure exactly the accountability-related variance that the current qualitative data found to be of the greatest importance to management practice.

This multidimensional reading may also be supported beyond the UAE setting: a systematic review of the trust-in-AI literature based on the trust theory finds that transparency, accountability, and reliability are independent rather than interchangeable indicators of trust, and that considering them as an undifferentiated construct blur which lever that actually is a driver of user confidence in high-stakes applications (Regona et al., 2026). The externality of the security domain to experimental evidence supports the special role of accountability-oriented trust: a study of human interaction in AI-aided decision-making finds that the availability of a visible, responsible human reviewer influences trust and perceived fairness in independent relation to the underlying accuracy of the AI system, which resembles how the current participants differentiated between technical confidence in AI outputs and confidence in the accountability structures that surrounded the same outputs (Choung et al., 2025).

Artificial intelligence (AI) Governance in the UAE Security Sector.

The results of governance directly relate to the overall question of how centralised, high-speed adoption governance regimes like the one of the UAE copes with the trade-off between the speed of deployment and institutional preparedness. The strategic investment by the UAE in modernisation based on AI has clearly led to faster deployment across sectors (al Sulaity et al., 2025; Iqbal et al., 2026), though the current results reveal that it has not been accompanied by

an equivalent acceleration in the formalisation of sector-specific governance to be applied in security implementation, although recent findings indicate that this has not been coupled with a commensurate acceleration of the process in which sector specific governance towards the application of security has been formalised. It is lagging in contrast to more developed forms of governance emphasis as seen in municipalities and healthcare, areas where greater mandates of public transparency have historically motivated more rapid governance formalisation (Alyileili & Opoku, 2025b, 2026). This imbalance cannot be interpreted merely as an oversight of policy based on the need to expand current governance templates to the security sector since the current results indicate that security-sector AI governance has its own unique demands, especially of dynamically calibrated and not fixed human-oversight forms, which current templates designed to address more public-facing service settings may not meet. It implies that UAE AI governance, to be effective in the security context, must be elaborated sector-specifically, and not uniformly extended across cross-sectoral frameworks, a fact that directly relates to national governance development currently since the comparative international experience shows responsible-AI governance frameworks to be most effective when tailored to the risk and accountability profile of the domain in which they are implemented (Mohammed et al., 2025). This tendency of rapid capability implementation ahead of formal accountability systems is not, however, specific to the UAE or the security industry: a qualitative survey of healthcare workers in a UK NHS Trust also revealed that the question of accountability, transparency, and bias issues remained even in the context of AI tools already integrated into the daily work process, specifically because the organisational governance frameworks have not kept up with the underlying technology (Nouis et al., 2025). This UK discovery of healthcare convergence with the current UAE security-sector outcomes in two highly dissimilar institutional and risk settings supports the argument that the governance-lag phenomenon observed here is indicative of a broader property of the fast diffusion of AI, as opposed to a jurisdiction phenomenon.

Conclusion

This paper aimed to explore how cybersecurity vulnerability and trust issues inform AI management practice in the UAE security field, and the key conclusion to be drawn here is that the two barriers do not exist as distinct management issues but are constitutive components of one underlying issue: the lack of strong and dynamically tuned accountability frameworks that can make the AI systems behaviour interrogateable when it needs to be. AI-specific cybersecurity issues, like adversarial manipulation and data-integrity compromise, have consequences that are not only undone by an accountability gap but also exacerbated by such gaps, as it leaves practitioners and oversight bodies without a resolved foundation on which to base responsibility when such vulnerabilities are exploited or when decisions influenced by AI fail in edge cases. Trust barriers, on their part, can be best perceived as not the lack of technological familiarity but a rational multiple dimensional reaction to this same accountability gap and a reaction that is in its turn fought by practitioners themselves via informal and formal measures that in their best attempts to implement are not a parallel move but rather an integrated one; one of security hardening and accountability-building.

The final claim of this paper is therefore that ongoing AI innovation in the UAE security sector cannot be managed as an independent, sequential project that is separate from the development of trustworthy governance structures. Present findings indicate that innovation without an adequate accountability system introduces the hidden nature of risk and unassigned liability which generate the trust deficits captured in this paper. When dealing with sustainable AI

management within this area, such that the formalisation of governance is considered to be a precondition and not a post-condition to the deployment, it is especially crucial to remember that failure in this sphere is irredeemable, and it is a matter of life and death.

This analysis gives rise to three recommendations. To start with, AI governance structures specific to the security sector must be developed by the UAE security organisations and their regulators in the context of a security sector with a risk and accountability profile that is tailored to the risk and accountability profile of the security sector, as opposed to extending governance templates created in the context of public-facing services where the consequence profile of failure is materially different. Second, organisations need to invest in cybersecurity practices, including zero-trust architectures expanded to AI components, which are explicitly intended to also have an accountability role, as the current evidence shows that strategies that fulfil both roles are perceived by the practitioners themselves to be more credible than parallel but separated ones. Thirdly, and most generally, sustainable AI governance in the UAE security sector needs to understand that trust is not a downstream effect of successful technical implementation which can be fostered by communication or training by itself, but rather a structural attribute of governance that needs to be built into AI systems and the institutions surrounding them in the first place.

Implications

Practical Implications

The most practical implication of these results as pertains to security organisations is that governance investment must focus on mechanisms that will not only bridge the accountability gap but also mitigate technical vulnerability, and not to consider cybersecurity hardening and trust-building as two budget lines. The strategies found in Theme 4, especially the application of zero-trust principles to AI elements and the creation of internal accountability and audit measures, would be tangible examples of this two-purpose investment, and organisations wishing to emulate them must not be tempted to see staff training as a one-dimensional solution that is not accompanied by a structural accountability reform, as the current evidence suggests that training without a structural accountability reform only creates the illusion of preparedness found in the organisational-readiness subtheme of Theme 3. In practice, it implies that security organisations must audit their AI governance artefacts not only to exist, that is, to have a policy document or a training programme, but to have substantive connection to an actual functioning, resourced accountability mechanism that can be put into effect in the face of real incidents.

Theoretical Contributions

This research article adds to AI management and cybersecurity literature by providing an empirically based explanation of cybersecurity and trust as interactive, as opposed to parallel, phenomena, which addresses the conceptual gap stated in the introduction directly. It also adds a multidimensional reconceptualisation of trust in the specific conditions of high-stakes operational settings, separating technological, human, and organisational dimensions in a way that has not been systematically expressed in the previous UAE-centric AI-adoption literature that has been inclined to assume trust to be a unitary acceptance construct. This work adds to qualitative insight into the issue of practitioner resistance to AI autonomy in particular, by providing evidence that practitioner resistance to AI autonomy is conditional and

accountability sensitive, rather than a generalised technological scepticism, which has the implication of how future qualitative research in this field should define its analytic terms.

Study Limitations

The results of this study are to be understood against a number of constraints. The sample, though stratified according to the relevant areas of professional interest, is confined in size, as is characteristic of qualitative exploratory research, and findings thereof are analytically, but not statistically generalisable. A sector-specific focus is also a methodological limitation of the study, and was an intentional approach, but restricts the degree to which the findings can immediately be applicable to non-security public-sector AI uses without further validation, as the evidence examined above indicates that risk-priority weighting may vary across sectors. Lastly, the qualitative design, although appropriate to the exploratory objectives of the present study, is not able to determine the relative scale or frequency of the patterns observed, which is why the mixed-method extension suggested below is being proposed.

Directions for Future Research

The above limitations and findings lead to three future research directions. First, a comparative qualitative analysis of security sectors in other Gulf states with centralised governance would serve to determine whether the trends observed in this case embody a UAE-specific governance set up or a more general aspect of fast, state-led AI diffusion. Second, mixed method designs that integrates the qualitative categories used in the current study, especially the multidimensional trust framework suggested above, with quantitative tools that can measure the relative importance of each trust dimension would enable testing the relative scale of the theoretical contribution to the present study. Third, longitudinal research that follows particular security organisations as they evolve through the governance-lag conditions that were reported in Theme 3 to more developed accountability structures would give direct evidence on whether the dual-purpose strategies found in Theme 4 actually bridge the accountability gap over time, rather than having to use practitioner retrospective evaluation of strategy success as the current cross-sectional design would need to do.

-
- Acknowledgements:** The authors would like to express their sincere gratitude to the Faculty of Management and Economic, Sultan Idris Education University, Malaysia, for providing the necessary resources and support throughout the course of this research. Special appreciation is extended to colleagues and peers who contributed valuable insights and constructive feedback, which greatly enhanced the quality of this paper.
- Funding Statement:** No Funding
- Conflict of Interest Statement:** The authors declare that there is no conflict of interest regarding the publication of this paper. All authors have contributed to this work and approved the final version of the manuscript for submission to the International Journal of Information System and Technology Management (JISTM).
- Ethics Statement:** Ethical processes were carefully followed considering that security-sector data is sensitive, and informed consent, confidentiality and anonymisation of organisational and individual identifiers, and special attention to ensure that no operational security information that could jeopardise organisational defences, was stored in the dataset or conveyed in findings. Ethical consent was taken before the data was gathered and participants informed of the boundaries of confidentiality that should be applied to a security-related study, including disclosures that might relate to systemic vulnerability.
- Author Contribution Statement:** Both authors contributed significantly to the development of this manuscript. Khalfan Abdulla was responsible for conceptualization, methodology, data collection, and drafting of the manuscript. Rafiduraida Abdulrahman contributed to the literature review, analysis and interpretation of results, and critical revision of the manuscript. Both authors read and approved the final version of the manuscript prior to submission.
-

References

- Abulbasher, A. (2026). Digital habitus and the sociotechnical transformation of everyday life in the United Arab Emirates: a hybrid framework of governmentality, surveillance capitalism, and Gulf state modernity. *Frontiers in Sociology*. <https://doi.org/10.3389/fsoc.2026.1766932>
- Al Badi, F. K., Alhosani, K. A., Jabeen, F., Shehzad, N., & Amann, W. C. (2022). Challenges of AI Adoption in the UAE Healthcare. *Vision*. <https://doi.org/10.1177/0972262920988398>
- Al Jaber, S. M., Patel, A., & Al-Masri, A. N. (2023). Object tracking and detection techniques under GANN threats: A systemic review. *Applied Soft Computing*. <https://doi.org/10.1016/j.asoc.2023.110224>
- Al Mazmi, F. I., & Radwan, A. F. (2026). Beyond Automation: Ethical and Professional Dimensions of Artificial Intelligence Adoption in UAE Government Communication. *Studies in Media and Communication*. <https://doi.org/10.11114/smc.v14i3.8324>
- Al Sulaity, I., Yourston, D., Khassawneh, O., & Darwish, T. K. (2025). Driving digital transformation in the oil and gas sector: unlocking AI 's potential in the UAE. *International Journal of Organizational Analysis*. <https://doi.org/10.1108/IJOA-01-2025-5145>
- Alattas Alhashmi, O. A., Abdullah, M. F., & Abdullah, R. S. (2023). A model of artificial intelligence in cyber security of scada to enhance public safety in UAE. *Journal of Theoretical and Applied Information Technology*.
- Alazzam, F. A. F., Gharaibeh, Z. I. Y., Jarah, B. A. F., Aljabali, A. M. A., & Al-Zaqeba, M. A. A. (2026). Legal and cybersecurity challenges of integrating artificial intelligence and the internet of things in financial institutions in the United Arab Emirates and Jordan. *International Journal of Data and Network Science*. <https://doi.org/10.5267/j.ijdns.2025.9.021>
- Alhashmi, S. F. S., Salloum, S. A., & Abdallah, S. (2020). Critical Success Factors for Implementing Artificial Intelligence (AI) Projects in Dubai Government United Arab Emirates (UAE) Health Sector: Applying the Extended Technology Acceptance Model (TAM). https://doi.org/10.1007/978-3-030-31129-2_36
- Almuraqab, N. A. S., Ateeq, A. A., Manoj Kumar, M. V., & Alfiras, M. I. I. (2025). Public attitudes toward secure AI enabled drone delivery for public services in the UAE. *Frontiers in Built Environment*. <https://doi.org/10.3389/fbuil.2025.1640830>
- Alrasheed, R. A., Al-Ameri, I. A., Benabdallah, A. K., Ahmed, H., & Sanad, H. (2025). Balancing Security and Morality: Arab Public Perspectives on Ethical Issues and Decision-Making in AI-Driven Security Policies. 2025 International Conference on Decision Aid Sciences and Applications, DASA 2025. <https://doi.org/10.1109/DASA68193.2025.11498959>
- Alshehhi, K., Cheaitou, A., & Rashid, H. S. J. (2024). Procurement of Artificial Intelligence Systems in UAE Public Sectors: An Interpretive Structural Modeling of Critical Success Factors. *Sustainability (Switzerland)*. <https://doi.org/10.3390/su16177724>
- Al-Skaf, S., Youssef, E. M., Habes, M., Alhumaid, K. F., & Salloum, S. A. (2021). The Acceptance of Social Media Sites: An Empirical Study Using PLS-SEM and ML Approaches. https://doi.org/10.1007/978-3-030-69717-4_52
- Alyileili, M., & Opoku, A. (2025a). Ethical AI integration in municipal self-service technologies: A case study from UAE public sector transformation. *Journal of Open Innovation: Technology, Market, and Complexity*, 11(4), 100648. <https://doi.org/10.1016/j.joitmc.2025.100648>

- Alyileili, M., & Opoku, A. (2025b). Ethical AI integration in municipal self-service technologies: A case study from UAE public sector transformation. *Journal of Open Innovation: Technology, Market, and Complexity*. <https://doi.org/10.1016/j.joitmc.2025.100648>
- Alyileili, M., & Opoku, A. (2026). Sustainable and Inclusive AI Governance in Municipal Self-Service Systems: Ethical, Smart-Government, and Generative AI Perspectives. *Sustainability (Switzerland)*. <https://doi.org/10.3390/su18020849>
- Ambali Parambil, M. M., Rustamov, J., Ahmed, S. G., Zaki, N. M., & Alnajjar, F. S. K. (2024). Integrating AI-based and conventional cybersecurity measures into online higher education settings: Challenges, opportunities, and prospects. *Computers and Education: Artificial Intelligence*. <https://doi.org/10.1016/j.caeai.2024.100327>
- Awad, A. I., Babu, A., Barka, E. S., & Shuaib, K. A. (2024). AI-powered biometrics for Internet of Things security: A review and future vision. *Journal of Information Security and Applications*. <https://doi.org/10.1016/j.jisa.2024.103748>
- Bahgat, A. (2025). AI governance and data privacy in cross-border contexts: A comparative analysis of regulatory frameworks. *Journal of Data Protection and Privacy*. <https://doi.org/10.69554/BOJL3033>
- Barbosa, G. N. N., Andreoni, M. A., & Mattos, D. M. F. (2025). Leveraging Zero Trust for Enhanced Security and Connectivity in Ad-Hoc Mesh Networks. 2025 12th IFIP International Conference on New Technologies, Mobility and Security, NTMS 2025. <https://doi.org/10.1109/NTMS65597.2025.11077020>
- Choung, H., David, P., & Mahmud, H. (2025). Fairness and trust in AI decision-making: The role of human involvement and outcome favorability. *International Journal of Human-Computer Interaction*. <https://doi.org/10.1080/10447318.2025.2576634>
- El Naggar, H., & Abdelrazik, H. (2024). Assessing community awareness for flood disasters in the UAE through human-centered design. *International Journal of Disaster Risk Reduction*. <https://doi.org/10.1016/j.ijdr.2024.104475>
- Hafiz, D. O., Hou, Q., & Zohdy, I. H. (2025). Advancing Sustainable Urban Mobility: Public Acceptance and Perceived Risks of Autonomous Vehicle Deployment in Dubai. *Sustainability (Switzerland)*. <https://doi.org/10.3390/su172411021>
- Khalil, E. M. A. S., & al-Ali, F. A. (2026). AI adoption in government public relations: Technology acceptance, social influence, and organizational dynamics in the UAE. *Public Relations Inquiry*. <https://doi.org/10.1177/2046147X251383106>
- Mohammed, A. Q., Khalifa, G. S. A., & Alhammedi, F. H. (2025). Evaluating Responsible AI Adaption: Ethics, Bias Mitigation, and Governance. 2025 10th International Conference on Information Technology Trends, ITT 2025. <https://doi.org/10.1109/ITT69610.2025.11352926>
- Nouis, S. C. E., Uren, V., & Jariwala, S. (2025). Evaluating accountability, transparency, and bias in AI-assisted healthcare decision-making: A qualitative study of healthcare professionals' perspectives in the UK. *BMC Medical Ethics*, 26, 89. <https://doi.org/10.1186/s12910-025-01243-z>
- Regona, M., Yigitcanlar, T., Hon, C., & Teo, M. (2026). Building trust in artificial intelligence: A systematic review through the lens of trust theory. *ACM Computing Surveys*, 58(9), Article 220. <https://doi.org/10.1145/3789256>
- Sebire, J., Stubbs, G., & Al Hanaee, M. (2025). Safety and security in Abu Dhabi: The impact of policing on public perceptions. *International Journal of Law, Crime and Justice*. <https://doi.org/10.1016/j.ijlcj.2025.100728>

ShwedeH, F. (2024). The Integration of Artificial Intelligence (AI) Into Decision Support Systems Within Higher Education Institutions. *Nanotechnology Perceptions*, 20(S5). <https://doi.org/10.62441/nano-ntp.v20is5.26>