



AI-Based Decision Support Systems and Administrative Efficiency in Saudi Public-Sector Organisations: The Mediating Role of Decision-Making Quality

Jamilah Abdullah Al Jarallah^{1,2}, Hanan Abdelrhman Al Edwan^{3*}, Faisal Sadun Alharbi⁴

¹College of Business, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

²Middle East Pioneering Company Limited (MEPCL). E-mail: jamilah.j.alqagtani@gmail.com, ORCID: <https://orcid.org/0009-0000-4071-2334>

³Department of Business Administration, Faculty of Economics, Al-Balqa Applied University, Jordan.

E-mail: aladwanable@gmail.com, ORCID: <https://orcid.org/0009-0003-4527-8962>

⁴Department of Human Resources Management and Organizational Studies, Essex Business School, University of Essex, United Kingdom. E-mail: fa24795@essex.ac.uk

*Corresponding author: Hanan Abdelrhman Al Edwan, aladwanable@gmail.com

Abstract

Purpose: Governments are increasingly adopting decision support systems (DSS) based on artificial intelligence (AI), yet little empirical research has examined how these systems affect administrative efficiency. Drawing on an integration of DeLone and McLean's information systems (IS) success model and task-technology fit (TTF) theory, this study examines how the information quality, system quality and task-technology fit of AI-based DSS (AI-DSS) relate to administrative efficiency through decision-making quality.

Design/methodology/approach: Survey data from 455 employees and managers in 22 Saudi public-sector organisations that use AI-DSS in their work were analysed using covariance-based structural equation modelling with robust maximum likelihood estimation, 10,000-resample bootstrapping for mediation analysis, and model comparison procedures.

Findings: All three AI-DSS characteristics were positively associated with decision-making quality ($R^2 = .545$), with task-technology fit the strongest predictor ($\beta = .407$). Decision-making quality was the strongest predictor of administrative efficiency ($\beta = .344$; $R^2 = .507$) and partially mediated all three relationships, accounting for 37-39% of each total effect. Task-technology fit and information quality also had direct positive associations with efficiency, whereas the direct association between system quality and efficiency was weak and not robust to alternative specifications.

Originality/value: This paper shifts the focus of research on public-sector AI from adoption to post-adoption value realisation. It identifies decision-making quality as a mechanism through which AI-DSS characteristics relate to administrative efficiency, and it distinguishes a judgement pathway, through which system characteristics improve decisions, from an execution pathway, through which fit and information quality streamline task performance directly.

Keywords: artificial intelligence; decision support systems; decision-making quality; administrative efficiency; task-technology fit; IS success model; public sector; Saudi Arabia

1. Introduction

Governments have moved beyond digitalising transactional processes to applying AI in decision-making. AI techniques such as predictive analytics, classification, prioritisation and recommendation are now embedded in case management, inspections, resource allocation and service delivery (Sun & Medaglia, 2019; Wirtz et al., 2019; van Noordt & Misuraca, 2022). Decisions are especially consequential in public administration because they allocate public resources, affect citizens' rights and must be justified to oversight bodies. Improving decision-making through AI is therefore a central challenge for both public administration and information systems (IS) research.

An AI-based decision support system (AI-DSS) is distinct from AI adoption in general and from the automation of decisions. An AI-DSS analyses data or generates predictions, classifications, risk assessments, recommendations or summaries that a human official uses when making a decision; it supports rather than replaces the decision-maker (Duan et al., 2019; Shrestha et al., 2019). It therefore differs both from traditional information systems, which store and retrieve data, and from automated systems, which act without human intervention (Bullock, 2019; Busuioc, 2021). The potential of such systems is considerable, since better-informed decisions should reduce delays, errors and inefficiencies (Desouza et al., 2020), but whether this potential is realised remains uncertain.

Deploying such systems does not guarantee improvement. Research on human-AI interaction shows that decision-makers may abandon algorithmic advice after observing it err (Dietvorst et al., 2015), may favour algorithmic over human advice and over-rely on it (Logg et al., 2019; Keding & Meissner, 2021), and may adhere to it selectively when it confirms their existing views (Alon-Barkat & Busuioc, 2023; Selten et al., 2023). Opacity can lead professionals to disengage from AI recommendations (Lebovitz et al., 2022), and collaboration with AI can improve individual performance while eroding aspects of human knowledge (Fügener et al., 2021). This suggests

that the value of an AI-DSS depends on characteristics that users can recognise and evaluate, namely information quality (the accuracy and relevance of outputs), system quality (reliability and responsiveness) and task-technology fit. However, these characteristics have rarely been examined as predictors of decision-making quality and efficiency.

The issue is particularly important in public administration, where decision-making is subject to legal, accountability and institutional pressures that apply to private decision-making to a lesser extent (Busuioc, 2021; Wirtz et al., 2022). AI-supported decisions must be explainable, the cost of errors is borne by citizens, and efficiency gains must not come at the expense of other public values (Bannister & Connolly, 2014; Andersen et al., 2016).

Saudi Arabia is a theoretically relevant setting for these questions. The National Strategy for Data and AI explicitly aims to integrate AI into government work to make government smarter and more efficient and to support decision-making (Saudi Data and AI Authority [SDAIA], 2020). The development and deployment of AI in the Saudi government have therefore been driven largely by policy, rather than emerging mainly from organisational needs as in many other settings. Case evidence shows that AI integration in Saudi public-sector organisations depends heavily on top-management attention, leadership and communication (Alshahrani et al., 2022, 2025).

Three gaps motivate this study. The first is a phenomenon gap. Recent research on AI in the public sector has concentrated on readiness, capabilities, governance and the antecedents of adoption (Mikalef et al., 2022; Madan & Ashok, 2023; Neumann et al., 2024), and literature reviews note a continuing lack of empirical evidence on the effects of AI use in public administration (Zuiderwijk et al., 2021; Medaglia et al., 2023; Alshahrani et al., 2024). Little attention has been paid to AI-DSS at the post-adoption stage, where value is created or destroyed (Jasperson et al., 2005; Burton-Jones & Straub, 2006).

The second is a mechanism gap. Repeated calls to “open the black box” in IS research have emphasised the need to identify the intermediate mechanisms that link technology to performance (Melville et al., 2004), and decision effectiveness is one such mechanism (Cao et al., 2015; Ghasemaghahi et al., 2018). The IS success model, however, treats individual and organisational impacts as broad categories without specifying the mechanisms that connect them (DeLone & McLean, 2003; Petter et al., 2008).

The third is a context gap. Research on AI in the Saudi public sector has focused mainly on assimilation, leadership and implementation (Alshahrani et al., 2022, 2025), and evidence on the outcomes of AI-DSS use by Saudi public-sector employees remains scarce.

This study addresses these gaps by integrating the IS success model (DeLone & McLean, 1992, 2003) with task-technology fit (TTF) theory (Goodhue & Thompson, 1995). Four research questions are posed. RQ1: How are the information quality, system quality and task-technology fit of AI-DSS related to decision-making quality in Saudi public-sector organisations? RQ2: How is decision-making quality related to administrative efficiency? RQ3: Do information quality, system quality and task-technology fit have direct relationships with administrative efficiency? RQ4: Does decision-making quality mediate the relationships between AI-DSS characteristics and administrative efficiency? The questions are answered using survey data from 455 employees and managers in Saudi public-sector organisations that use an AI-DSS in their work.

The study makes four contributions. First, it shifts research on public-sector AI from adoption to post-adoption value by examining the decision and efficiency outcomes reported by actual users. Second, it integrates the IS success model and TTF theory around an outcome rather than around use, showing that quality-based and fit-based characteristics relate to efficiency in different ways. Third, it specifies and tests decision-making quality as the mechanism linking AI-DSS characteristics to administrative efficiency. Fourth, it provides confirmatory evidence from the Saudi public sector, where policy-driven AI adoption makes the gap between deployment and performance especially salient. The remainder of the paper reviews the theoretical background (Section 2), develops the hypotheses (Section 3), describes the methodology (Section 4), presents the results (Section 5), discusses the findings and their implications (Sections 6-9), notes the limitations (Section 10) and concludes (Section 11).

2. Literature Review and Theoretical Background

2.1 AI-based decision support systems

Decision support systems (DSS) are computer-based systems that help decision-makers use data and models to address semi-structured and unstructured problems. Traditional DSS combined database, model and interface components that supported human decision-making without taking it over. Intelligent DSS added knowledge-based and rule-based components capable of reasoning with expert knowledge. Contemporary AI-DSS use machine learning, natural language processing and generative models to perform tasks beyond the reach of earlier generations, such as prediction, classification, recommendation and prioritisation, anomaly detection, scenario evaluation and the summarisation of unstructured data (Duan et al., 2019; Dwivedi et al., 2021).

AI's contribution to DSS is therefore not only speed but also the provision of analytical information that decision-makers would otherwise have to generate themselves. At the same time, AI has features that make it harder to link technology use to added value: AI systems learn and evolve, their reasoning is often difficult to interpret, and they operate with a degree of autonomy (Berente et al., 2021). A substantial body of research argues that organisational AI should be understood as augmentation rather than automation, with humans drawing on algorithms to improve

their decisions without surrendering their judgement (Jarrahi, 2018; Raisch & Krakowski, 2021). Reviews and studies of human-AI collaboration in decision-making reach the same conclusion and stress the importance of integrating algorithmic outputs into human decision processes (Trunk et al., 2020; Haesevoets et al., 2021). This study adopts the augmentation perspective: officials continue to make decisions, supported by AI-DSS.

2.2 AI and decision-making in public-sector organisations

AI can benefit public organisations in several ways: faster handling of routine cases, more consistent treatment of similar cases, early detection of risks and better targeting of inspection and service resources (Wirtz et al., 2019; Desouza et al., 2020). Mapping studies show that governments use AI for prediction, classification and recommendation across many policy domains (van Noordt & Misuraca, 2022), and public servants often view AI as a way to improve their performance (Ahn & Chen, 2022).

The literature is equally clear about the challenges. Research on algorithmic decision-making in government highlights problems of data quality, transparency, accountability and the alignment of AI with organisational processes (Sun & Medaglia, 2019; Kuziemski & Misuraca, 2020; Busuioc, 2021). AI also reshapes administrative discretion, shifting some judgements from street-level bureaucrats to system designers and data (Bullock, 2019). Experimental studies show that civil servants do not follow algorithmic advice blindly; rather, they adhere to it selectively, particularly when it aligns with existing stereotypes or their professional judgement (Alon-Barkat & Busuioc, 2023; Selten et al., 2023). Explanation and transparency can shape the perceived trustworthiness of AI-driven decisions (Grimmelikhuijsen, 2023), and the effect of explainable AI on decision accuracy depends on the decision-maker's experience (Janssen et al., 2022). Trust in AI depends on its tangibility, transparency and reliability (Glikson & Woolley, 2020), and algorithm aversion remains a concern in augmented decision-making (Burton et al., 2020).

Two implications follow for this study. First, the usefulness of AI for decision-making depends heavily on officials' perceptions of the system, based on the quality of its outputs, its performance and its suitability for the task. Second, trust, transparency and governance are relevant contextual conditions, but they are not characteristics of the system itself, which is the focus of this research; they are therefore treated here as contextual literature. Table 1 summarises representative studies in this space.

Table 1. Representative Studies on AI in Public Administration and Their Relation to the Present Study

Study	Design and context	Focal question	Gap relative to this study
Sun & Medaglia (2019)	Qualitative; public healthcare, China	How stakeholders frame AI challenges	No measurement of decision or efficiency outcomes
Zuiderwijk et al. (2021)	Systematic review	Implications of AI for public governance	Notes the scarcity of empirical impact studies
Mikalef et al. (2022)	Survey; European municipalities	Determinants of AI capabilities	Organisational capability, not use outcomes
Ahn & Chen (2022)	Survey; government employees	Willingness to use AI in government	Intention to use, not realised outcomes
Alshahrani et al. (2022)	Case study; Saudi public sector	Managerial attention and AI assimilation	Assimilation process; no user-level outcomes
Madan & Ashok (2023)	Systematic review	AI adoption and diffusion in public administration	Adoption focus; post-adoption value under-studied
Alon-Barkat & Busuioc (2023)	Experiments; civil servants	Automation bias and selective adherence	Decision behaviour, not system characteristics
Selten et al. (2023)	Experiment; police officers	Trust in AI recommendations	Trust as outcome; efficiency not examined
Alshahrani et al. (2024)	Systematic review	AI and decision-making in government functions	Calls for empirical evidence on outcomes
Neumann et al. (2024)	Comparative case study	Drivers of AI adoption in public organisations	Adoption stage; no outcome measurement
Alshahrani et al. (2025)	Case studies; Saudi public sector	Leadership and communication in AI assimilation	Assimilation; no decision or efficiency outcomes

2.3 Digital government in Saudi Arabia and AI

Saudi Arabia's digital government agenda has shifted from digital service delivery towards data- and AI-enabled government. The Saudi Data and AI Authority (SDAIA) was established to lead this agenda, and government is identified as a priority sector in the National Strategy for Data and AI, which counts whole-of-government analytics

and data-driven decision-making among its key aims (SDAIA, 2020). Most scholarly work on this transition has focused on the adoption and integration of AI.

Drawing on the attention-based view, Alshahrani et al. (2022) found that AI assimilation in Saudi public organisations depends on how top management allocates its attention. A subsequent multiple-case study identified the role of leadership and communication in AI assimilation in Saudi Arabia (Alshahrani et al., 2025), and a systematic review examined the opportunities and challenges of using AI in decision-making across government functions (Alshahrani et al., 2024). These studies provide a clear understanding of how Saudi government organisations adopt AI, but none examines the consequences of adoption for decision-making and administrative operations.

2.4 DeLone and McLean's IS Success Model

DeLone and McLean's (1992) IS success model organised diverse IS success research into six dimensions: system quality, information quality, use, user satisfaction, individual impact and organisational impact. Their updated model added service quality and combined individual and organisational impacts into net benefits (DeLone & McLean, 2003). The quality dimensions have been operationalised as object-based beliefs that shape users' behavioural beliefs about a system (Wixom & Todd, 2005), and their determinants have been studied extensively (Nelson et al., 2005). Gable et al. (2008) reframed the model as an IS-impact measurement model in which information quality and system quality capture the quality of the artefact, while individual and organisational impact capture its effects.

The model has also attracted criticisms that are relevant to this research. Seddon (1997) argued that the original model conflated process and variance explanations and that "use" was poorly defined. Reviews show that the model's relationships vary with the level of analysis, the type of system and the voluntariness of use (Sabherwal et al., 2006; Petter et al., 2008; Jeyaraj, 2020). The link between information quality and individual impact is relatively consistent, whereas the link between system quality and individual impact is weaker and more context-dependent (Petter et al., 2008). Most importantly, the model does not specify how individual impact arises: "individual impact" encompasses decision quality, productivity, learning and job performance, and research on the model's antecedents has neglected the process through which quality translates into individual benefits (Petter et al., 2013).

2.5 Task-technology fit theory

TTF theory holds that information technology is more likely to improve performance when its capabilities match the tasks that users perform (Goodhue & Thompson, 1995). In the technology-to-performance chain, fit affects performance both directly and indirectly through utilisation. Goodhue (1998) developed and validated user evaluations of fit, and Ziguers and Buckland (1998) formalised the theory by specifying ideal task-technology fit profiles for group support systems.

Three points shape how the theory is applied here. First, prior research has generally used fit to explain utilisation, although the theory originally emphasised performance (Goodhue & Thompson, 1995). Second, fit is not a monotonic function of capability, because outcomes can deteriorate when technological capabilities exceed task needs (Aljukhadar et al., 2014); fit therefore carries information that quality scores alone cannot capture. Third, fit is usually measured through user perceptions, which risks conceptual overlap with perceived quality (Goodhue, 1998). This study addresses the third issue by distinguishing fit from quality empirically.

Fit is central to AI-based decision support. The sophistication of an AI technology does not by itself guarantee benefits. AI is better suited to complex, data-rich problems than to highly uncertain ones (Jarrahi, 2018), and different types of decisions call for different divisions of labour between humans and machines (Shrestha et al., 2019). An AI-DSS that produces good risk indicators will therefore be of limited use to an official whose task is planning, and vice versa.

2.6 Decision-making quality

Decision-making quality (DMQ) refers to the soundness, accuracy, consistency and dependability of the decisions a decision-maker makes. In this paper, DMQ is distinguished from efficiency, because a decision process can be efficient yet produce poor decisions, and vice versa. Decision support theory links decision quality to the quality of available information and to the decision-maker's ability to use it (Raghunathan, 1999), as well as to changes in the cognitive effort required by different decision strategies (Todd & Benbasat, 1999).

Empirical research in organisational settings supports a link between IS capabilities and decision quality. Greater business intelligence maturity improves information quality and, in turn, analytical decision-making (Popovič et al., 2012); data analytics competency improves decision quality and decision efficiency (Ghasemaghahi et al., 2018); business analytics improves decision-making effectiveness through information processing (Cao et al., 2015); and enterprise systems are associated with higher decision-making quality (Ouidad et al., 2021). Within the impact chain of the IS success model (DeLone & McLean, 2003; Gable et al., 2008), these results suggest that decision-making quality is the individual-level impact most proximal to the purpose of an AI-DSS, which is decision support. DMQ is therefore modelled here as the intermediary mechanism.

2.7 Administrative efficiency

Administrative efficiency (AE) refers to the extent to which AI-DSS use saves time, increases productivity and work-processing capacity, reduces unnecessary procedural effort, uses resources efficiently and minimises avoidable rework. Public-sector performance is multidimensional, and efficiency, the ratio of administrative

outputs to inputs, is conceptually distinct from effectiveness, equity, service quality and other criteria (Boyne, 2003; Andersen et al., 2016). Reviews of the public value of e-government identify administrative efficiency as one of its main value dimensions (Twizeyimana & Andersson, 2019), although efficiency is only one of many values that technology may serve or undermine (Bannister & Connolly, 2014).

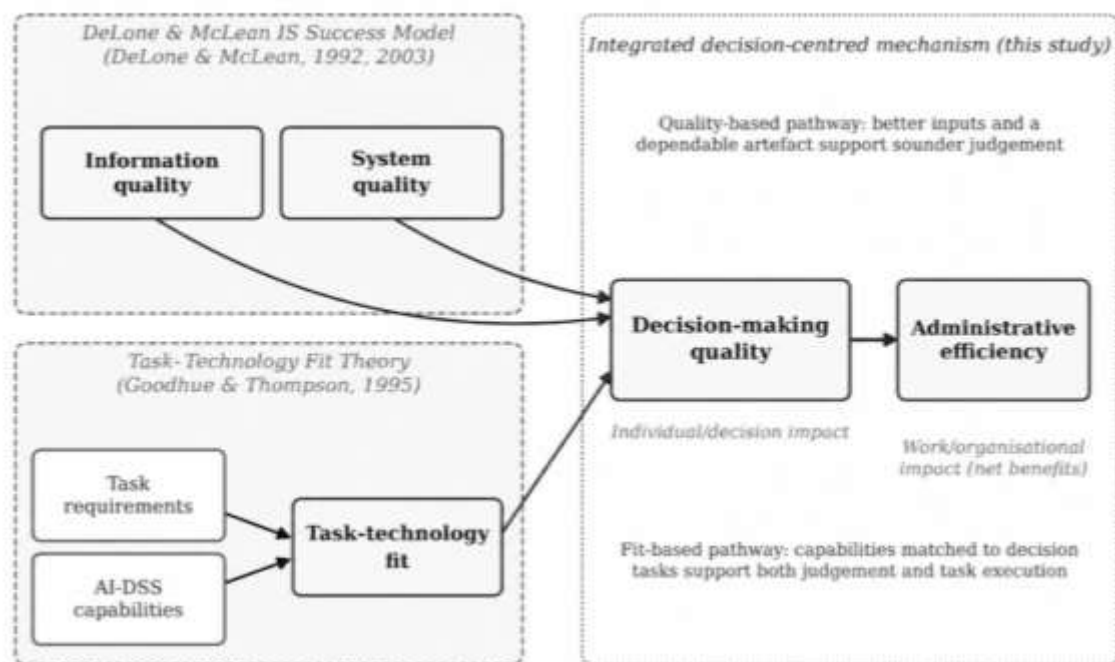
The construct is defined at the level of the individual and their immediate work group, consistent with the work-impact approach in IS research (Torkzadeh & Doll, 1999; Gable et al., 2008). It is narrower than organisational effectiveness or service quality, since an AI-DSS may speed up work without improving service quality, and vice versa. Research on IT value creation in the public sector suggests that IT resources create value indirectly through organisational capabilities (Pang et al., 2014), while digital-era governance theory identifies the digitisation and reintegration of administrative processes as a source of efficiency in its own right (Dunleavy et al., 2006). Together, these arguments imply that efficiency may arise both from better decisions and from improved task execution.

2.8 Synthesis and research gap

The two theories are complementary rather than competing. The IS success model directs attention to the properties of the artefact: the information the system provides (information quality) and how well it functions (system quality). TTF theory directs attention to the relationship between the artefact and the work: whether what the system provides meets the requirements of the task. A system may provide timely and accurate information that is nevertheless poorly suited to the decisions a manager faces; conversely, a modest system may suit simple decisions well. Neither theory alone explains how both types of characteristics create value. Integrating them around decision-making quality addresses this limitation (Figure 1).

This integration makes it possible to test whether quality-related and fit-related characteristics of AI-DSS create value in the same way, through better decisions, or whether fit also affects the efficiency of task performance directly. As the review above shows, this question remains unanswered for AI-DSS in public administration: studies of AI in public administration address adoption, governance and trust; IS success research has not addressed AI-DSS; and Saudi studies examine assimilation rather than impact. Table 2 defines the five constructs of the integrated model.

Figure 1. Theoretical Framework: Integration of the IS Success Model and Task-Technology Fit Theory



Note. The dashed panels represent the source theories, and the dotted panel represents the integrated decision-centred model. Direct paths from the three AI-DSS characteristics to administrative efficiency were also tested (Figure 2).

Table 2. Construct Definitions and Theoretical Foundations

Construct	Definition in this study	Theory	Measurement source	Items
Information quality (IQ)	Perceived accuracy, timeliness, usefulness, clarity and decision relevance of the outputs produced by the AI-DSS	IS Success Model	DeLone & McLean (2003); adapted to AI-DSS	5

Construct	Definition in this study	Theory	Measurement source	Items
System quality (SQ)	Perceived availability, responsiveness, ease of use, functionality and reliability of the AI-DSS as a technical artefact	IS Success Model	DeLone & McLean (2003); adapted to AI-DSS	5
Task-technology fit (TTF)	Degree to which the functions and capabilities of the AI-DSS correspond to the requirements of the respondent's decision tasks	TTF theory	Goodhue & Thompson (1995); Goodhue (1998)	4
Decision-making quality (DMQ)	Perceived soundness, accuracy, consistency and dependability of decisions made with AI-DSS support	Decision-centred mechanism	Adapted from Ouidad et al. (2021)	4
Administrative efficiency (AE)	Perceived contribution of the AI-DSS to time savings, productivity and capacity, and the elimination of unnecessary effort and redundant procedures	IS impact / net benefits	Torkzadeh & Doll (1999) for AE1-AE3; context-specific AE4-AE6 after Gable et al. (2008)	6

Note. All constructs are reflective. Respondents answered every item with reference to one primary AI-DSS.

3. Hypotheses Development and Conceptual Model

3.1 Information quality and decision-making quality

Decisions depend heavily on the quality of the information available. Raghunathan (1999) showed that improvements in information quality raise decision quality, particularly when decision-makers understand the problem context. Organisational studies support this link: business intelligence maturity improves decision-making through information quality (Popovič et al., 2012), data quality is a core element of the analytics competency that improves decision quality (Ghasemaghaei et al., 2018), and information processing mediates the relationship between business analytics and decision-making effectiveness (Cao et al., 2015). In IS success research, information quality is among the most consistent predictors of individual impact (Petter et al., 2008; Jeyaraj, 2020).

The argument is even clearer for AI-DSS, whose distinctive contribution lies in its outputs: predictions, risk scores and recommendations that officials would otherwise have to produce themselves (Duan et al., 2019). High-quality outputs provide a more complete evidence base for decisions, help identify options and risks, and promote consistency across similar cases. Two caveats apply. Perceived information quality is not the same as actual model accuracy and may reflect the congruence between system outputs and the official's own views (Selten et al., 2023), and high-quality outputs may encourage over-reliance on the tool (Keding & Meissner, 2021).

H1. Information quality is positively associated with decision-making quality.

3.2 System quality and decision-making quality

System quality concerns the AI-DSS as a technical artefact, including its availability, response time, ease of use and reliability (DeLone & McLean, 2003; Gable et al., 2008). System quality affects decision-making through cognitive effort (Todd & Benbasat, 1999): decision-makers trade off accuracy against effort and adopt the strategies that their decision aids make easy. A decision aid that is unreliable or unresponsive raises the effort cost of using it and encourages decision-makers to bypass it. Reliability is also an important antecedent of trust in AI (Glikson & Woolley, 2020), and a lack of trust may lead decision-makers to disregard a decision support tool altogether. Empirically, system quality shapes users' perceptions of a system (Wixom & Todd, 2005; Sabherwal et al., 2006) and has been linked to outcomes in public taxation (Floropoulos et al., 2010) and in organisational settings (Gorla et al., 2010).

The literature nevertheless suggests that this effect is likely to be weaker than that of information quality, because evidence on the relationship between system quality and individual impact is mixed (Petter et al., 2008; Jeyaraj, 2020) and system quality functions mainly as an enabling condition.

H2. System quality is positively associated with decision-making quality.

3.3 Task-technology fit and decision-making quality

According to TTF theory, technology improves performance when its capabilities fit the requirements of the task (Goodhue & Thompson, 1995). In decision-making, fit means that the AI-DSS provides the kind of support the task requires, such as risk classification for triage tasks or scenario forecasting for planning tasks. When fit is high, outputs can be applied directly to the task; when fit is low, outputs must be translated, supplemented or discarded. Fit theory for group support systems makes the same argument (Zigurs & Buckland, 1998), and research on algorithmic decision support suggests that fit depends on the complexity, uncertainty and ambiguity of the decision task (Jarrahi, 2018; Shrestha et al., 2019). There is also evidence that surplus capabilities can be detrimental (Aljukhadar et al., 2014).

H3. Task-technology fit is positively associated with decision-making quality.

3.4 Decision-making quality and administrative efficiency

In administration, the quality of a decision directly affects resource use further down the chain. Sound decisions lead to fewer reconsiderations and appeals. They also reduce coordination costs, because similar cases are decided in similar ways, saving the time otherwise spent resolving poor or disputed decisions. Research on analytics indicates that decision capabilities improve not only decision quality but also decision efficiency (Ghasemaghaei et al., 2018), and that decision-making effectiveness is the mechanism through which analytics improves organisational performance (Cao et al., 2015).

In public organisations, however, IT creates value through enhanced organisational capabilities rather than through technology alone (Pang et al., 2014), and public service improvement stems largely from management and process (Boyne, 2003). Better decisions may sometimes take more time, for example when AI outputs flag new risks that require review, and technology can even reinforce bureaucracy (Cordella & Tempini, 2015). Nevertheless, the overall expectation remains positive, because the rework costs of poor decisions generally exceed the time spent making better ones.

H4. Decision-making quality is positively associated with administrative efficiency.

3.5 Direct relationships between AI-DSS characteristics and administrative efficiency

Not all efficiency effects of AI-DSS need to operate through better decisions. Digitised and automated administrative procedures are a distinct source of efficiency in digital-era governance (Dunleavy et al., 2006). An AI-DSS can generate efficiency simply by compiling information that officials would otherwise gather manually and by prioritising cases so that effort is directed where it is needed. Both theoretical frameworks anticipate direct effects. The IS success model posits direct effects of system and information quality on impacts (DeLone & McLean, 2003; Gorla et al., 2010), and TTF theory posits direct effects of fit on performance (Goodhue & Thompson, 1995). Fit in particular can be expected to affect efficiency directly, because a system that fits a task reduces the work required to perform that task, regardless of whether the decision itself improves (Todd & Benbasat, 1999). Retaining all three direct paths is also methodologically necessary to distinguish partial from full mediation.

H5. Information quality is positively associated with administrative efficiency.

H6. System quality is positively associated with administrative efficiency.

H7. Task-technology fit is positively associated with administrative efficiency.

3.6 The mediating role of decision-making quality

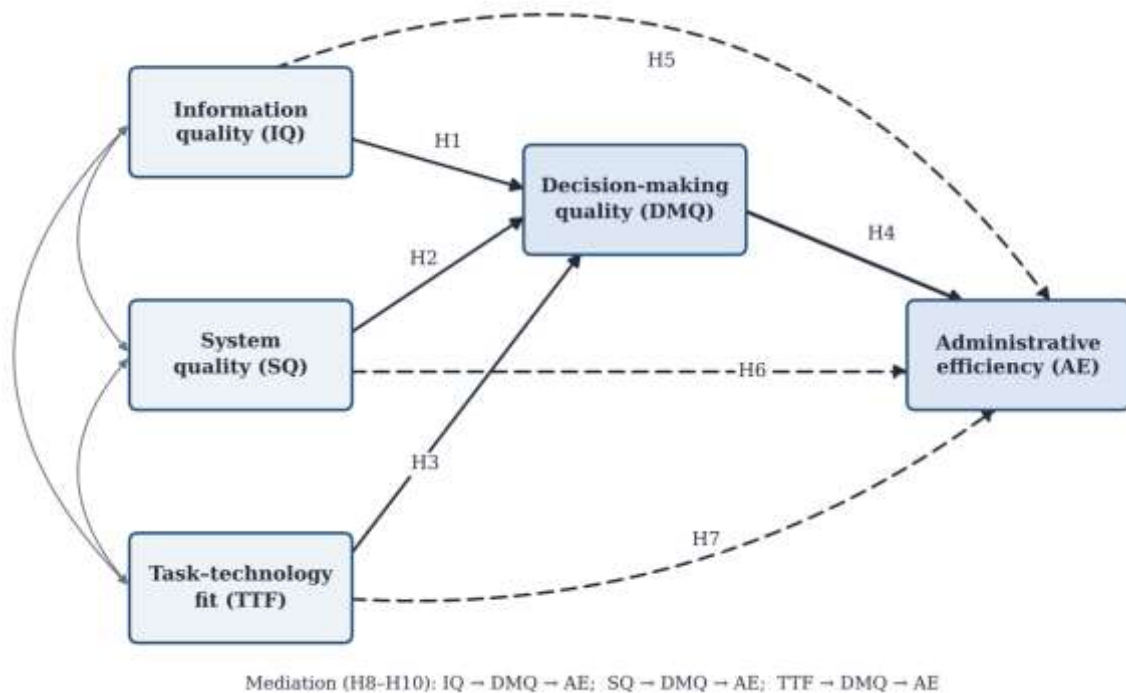
Together, these arguments imply mediation, and the mediating role of decision-making quality is theoretically grounded rather than incidental. The impact chain of the IS success model links artefact quality to individual impact, and individual impact to organisational impact (DeLone & McLean, 2003; Gable et al., 2008). Research on IT business value holds that the effects of technology are mediated by business processes rather than direct (Melville et al., 2004), and analytics research identifies decision-making effectiveness as one such process (Cao et al., 2015; Ghasemaghaei et al., 2018). For an AI-DSS, decision-making quality is the most proximal individual impact, because the primary purpose of the tool is to support decisions. Information quality, system reliability and task suitability can improve the quality of officials' decisions, and administrative efficiency then follows from fewer errors and less rework.

H8. Decision-making quality mediates the relationship between information quality and administrative efficiency.

H9. Decision-making quality mediates the relationship between system quality and administrative efficiency.

H10. Decision-making quality mediates the relationship between task-technology fit and administrative efficiency.

Figure 2 presents the conceptual model. Information quality, system quality and task-technology fit are specified as correlated exogenous constructs, reflecting that they are related facets of one user's experience of the same AI-DSS. Decision-making quality is the mediator, and administrative efficiency is the dependent variable.

Figure 2. Conceptual Model and Hypotheses


Note. Solid arrows denote the decision-quality pathway (H1-H4); dashed arrows denote direct efficiency paths (H5-H7). H8-H10 are the specific indirect effects through DMQ. Curved grey double-headed arrows indicate covariances among the exogenous constructs.

4. Methodology

This section describes the research design, instrument, sampling, data collection and analysis procedures used to test the research model. The analytical procedures were specified before data collection and followed the two-step approach to covariance-based structural equation modelling (CB-SEM), in which the measurement model is assessed before the structural relationships are tested (Anderson & Gerbing, 1988; Gaskin et al., 2025).

4.1 Research design

The study used a quantitative, explanatory survey design with a deductive approach. Relationships among the five latent constructs were hypothesised a priori from the IS success model (DeLone & McLean, 1992, 2003) and TTF theory (Goodhue & Thompson, 1995) and tested using latent variable modelling. A survey design is appropriate because information quality, system quality, task-technology fit, decision-making quality and administrative efficiency are experiential constructs best captured through users' own assessments.

CB-SEM was preferred to variance-based methods because the research is confirmatory: the aim was to test whether a theoretically specified measurement model and a network of direct and indirect relationships were consistent with the data while accounting for measurement error. Because all variables were measured at a single time point without experimental manipulation, the design cannot establish causal relationships.

4.2 Research model

The research model tested is the one developed in Sections 2 and 3 (Figures 1 and 2). IQ, SQ and TTF are correlated exogenous constructs, DMQ is the mediator, and AE is the dependent variable. The seven direct paths (H1-H7) and three specific indirect effects (H8-H10) are estimated simultaneously, and the direct paths from IQ, SQ and TTF to AE are retained so that partial and full mediation can be compared.

4.3 Research setting and unit of analysis

The empirical context is the public sector of the Kingdom of Saudi Arabia, including ministries, government agencies, municipalities, public universities, public health organisations, judicial and security bodies, and other public agencies in which employees use AI-supported systems for decision-making tasks. The unit of analysis is the individual employee or manager. All five constructs are measured with reference to the respondent's own experience of one specific AI-DSS. Individual responses are not aggregated to the organisational level, because the number of respondents per organisation is too small for this purpose.

An AI-DSS was defined as any system that uses AI to analyse information or to predict, recommend, classify, prioritise, assess risk, summarise or provide other analytical inputs used in work-related decisions. Only systems that provide analysis or predictions were included; systems that only store data or execute transactions were excluded. Respondents were asked to answer all questions with reference to a single AI-DSS that they use directly or rely on, to avoid confounding judgements about several systems of differing quality.

4.4 Population, eligibility and sampling

The target population comprised employees and professionals currently working in Saudi public-sector organisations who had experience of AI-based decision support. The sampling frame was defined behaviourally rather than by employment alone. To be eligible, participants had to (a) be currently employed in a Saudi public-sector organisation, (b) hold a position in which they make, contribute to or advise on administrative, operational, regulatory or strategic decisions, and (c) have used an AI system to support a decision within the previous six months.

Because no national register of AI-DSS users in the Saudi public sector exists, the study used non-probability purposive (eligibility-based) sampling. Respondents were recruited across several organisation types, managerial and decision-making levels and AI-DSS categories, rather than from a single organisation. Because this is not a probability sample, sample percentages cannot be interpreted as estimates of the prevalence of AI-DSS use in the Saudi public-sector workforce.

4.5 Sample size

Sample size was determined through a priori power analysis for SEM rather than a respondent-to-item ratio. Westland (2010) showed that the sample size required in SEM depends non-linearly on model complexity and on the minimum effect size to be detected, and simulation studies have shown that rules of thumb based on fixed ratios can substantially over- or underestimate the sample needed (Wolf et al., 2013). With 24 indicators, five latent variables, $\alpha = .05$, desired power of .95 and a minimum detectable effect size of .25, the minimum sample required for the model structure was 92 and the minimum required to detect the specified effect was 341 (Table 3). Accordingly, 341 valid responses was set as the formal minimum, 400 usable cases as the target sample and 450-500 completed surveys as the recruitment goal.

Table 3. A Priori Sample-Size Determination

Input / result	Value
Observed indicators (j)	24
Latent constructs (k)	5
Significance level (α)	.05
Desired statistical power	.95
Anticipated minimum effect size	.25
Minimum sample for model structure	92
Minimum sample to detect the specified effect	341
A priori analytical minimum	341
Planned usable sample	400
Achieved analytic sample	455

Note. Lower bounds follow Westland (2010). For the achieved sample, the power of the test of close fit (H_0 : RMSEA = .05 vs. H_a : RMSEA = .08, $df = 242$) exceeded .99 (MacCallum et al., 1996).

The achieved analytic sample of 455 exceeded both the formal minimum and the target sample. For the final model ($df = 242$), power to distinguish close fit (RMSEA = .05) from poor fit (RMSEA = .08) exceeded .99 (MacCallum et al., 1996). The sample was therefore sufficient for the structural and mediation analyses.

4.6 Measurement instrument

Data were collected using a bilingual Arabic-English questionnaire comprising one consent item, three screening items on the inclusion criteria (S1-S3), three items on AI-DSS use characteristics (duration of use, frequency of use and system type; S4-S6), 24 items measuring the five reflective constructs, and eight background items (D1-D8). The 24 construct items were adapted from established IS success, task-technology fit and IT impact measures and tailored to AI-DSS use in the Saudi public sector (see Table 2 in Section 2.8 and Table 4). All 24 items used a seven-point Likert scale (1 = strongly disagree to 7 = strongly agree).

Table 4. Questionnaire Items and Item-Source Mapping

Code	Item (English version)	Source basis	Status
IQ1	The information or outputs provided by the system are accurate.	DeLone & McLean (2003)	Adapted
IQ2	The system provides information at the time it is needed to support my decisions.	DeLone & McLean (2003)	Adapted
IQ3	The information provided by the system is useful for decisions related to my work.	DeLone & McLean (2003)	Adapted

Code	Item (English version)	Source basis	Status
IQ4	The output from the system is easily comprehensible.	DeLone & McLean (2003)	Adapted
IQ5	The information provided by the system is directly relevant to the decisions I need to make.	DeLone & McLean (2003)	Adapted
SQ1	The system is available and works when I need it.	DeLone & McLean (2003)	Adapted
SQ2	The system responds quickly enough to support the decision-making process.	DeLone & McLean (2003)	Adapted
SQ3	Using the system for decision-related tasks is easy for me.	DeLone & McLean (2003)	Adapted
SQ4	The system provides the functions and features needed to support decisions in my work.	DeLone & McLean (2003)	Adapted
SQ5	The system operates reliably and consistently during use.	DeLone & McLean (2003)	Adapted
TTF1	The system's functions are sufficient to support the decision-making tasks I perform.	Goodhue & Thompson (1995); Goodhue (1998)	Adapted
TTF2	The system's functions are appropriate for the nature of the decisions required in my work.	Goodhue & Thompson (1995); Goodhue (1998)	Adapted
TTF3	The system's capabilities meet the requirements of my decision-making tasks.	Goodhue & Thompson (1995); Goodhue (1998)	Adapted
TTF4	Overall, the system fits well with the tasks for which I use it.	Goodhue & Thompson (1995); Goodhue (1998)	Adapted
DMQ1	The system helps me make decisions whose results are usually reasonable.	Ouiddad et al. (2021)	Adapted
DMQ2	It enables me to make decisions through proper evaluation of the information that is at hand.	Ouiddad et al. (2021)	Adapted
DMQ3	It is beneficial for me to make consistent decisions when faced with identical situations or problems.	Ouiddad et al. (2021)	Adapted
DMQ4	I can rely on the quality of the decisions I make with the support of this system.	Ouiddad et al. (2021)	Adapted
AE1	Using the system helps me save time when completing administrative work related to my responsibilities.	Torkzadeh & Doll (1999)	Adapted
AE2	Using the system increases my productivity in administrative work.	Torkzadeh & Doll (1999)	Adapted
AE3	Through use of the system I am able to get more done than otherwise possible.	Torkzadeh & Doll (1999)	Adapted
AE4	Use of the system will minimise the unnecessary efforts in conducting administrative procedures.	Gable et al. (2008) impact logic	Context-developed
AE5	The use of this system allows me to manage more work without any proportional amount of additional time being needed.	Gable et al. (2008) impact logic	Context-developed
AE6	Using the system reduces the need for avoidable rework or corrections.	Gable et al. (2008) impact logic	Context-developed

Note. In the data file the DMQ items are coded DQ1-DQ4. "Adapted" items were reworded for the AI-DSS/public-sector context; "context-developed" items extend the AE construct to process efficiency and were validated empirically in this study (Section 5.4).

Several design measures were taken to protect measurement integrity. Construct names were not shown to respondents, so that the hypothesised model could not easily be inferred. The 14 antecedent items (IQ, SQ and TTF) were presented in random order, followed by separate blocks of DMQ and AE items, to create a logical sequence from system experience to decision outcomes to work outcomes. Reverse-worded items were not used, because they can introduce method effects, comprehension difficulties and translation problems.

4.7 Translation and cross-language equivalence

The Arabic and English versions were treated as linguistic versions of the same instrument, and translation aimed at conceptual rather than literal equivalence. A bilingual translator familiar with IS and organisational terminology translated the English version into Arabic; a second bilingual translator, blind to the original wording, back-translated the Arabic version into English; and discrepancies were discussed and resolved by the research team (Brislin, 1970; Beaton et al., 2000). Particular attention was paid to terms whose meaning may vary across administrative contexts, such as “decision support”, “system outputs”, “administrative procedures”, “rework” and “decision task”.

4.8 Content validity

Before data collection, a panel of five experts in information systems, artificial intelligence, public administration, organisational management and quantitative measurement independently rated each item for relevance, clarity and contextual appropriateness. Content validity indices were calculated following Polit et al. (2007), with thresholds of $I-CVI \geq .78$ for items and $S-CVI/Ave \geq .90$ for the scale. Content validity was satisfactory: $I-CVI$ values ranged from .80 to 1.00 and $S-CVI/Ave$ was .94. Minor wording and clarity revisions were made in response to the experts' qualitative feedback, but no items were removed, as all met the threshold and were judged relevant to their constructs. The indices were used as guidelines rather than strict decision rules, and final decisions also took account of the experts' qualitative feedback and the need for adequate construct coverage (MacKenzie et al., 2011).

4.9 Cognitive pretesting and pilot study

Following the expert review, cognitive interviews were conducted with 10 individuals with characteristics similar to those of the target respondents. Participants were asked to explain how they understood each item and how they selected their answers. The interviews indicated that the questionnaire was clear, although the wording of some items was refined for clarity.

A pilot study was then conducted with 45 eligible AI-DSS users to evaluate the screening criteria, questionnaire flow, completion time, functioning of the response categories and internal consistency of the measures. Participants reported no difficulties completing the survey, the average completion time was approximately 10–12 minutes, and the response categories functioned as intended. All constructs showed satisfactory internal consistency (Cronbach's $\alpha > .70$). Minor changes to the wording and order of some items were made in light of the pilot results.

4.10 Data collection procedure

Data were collected online between 12 April and 28 May 2026 using Microsoft Forms. The survey link was distributed by email and WhatsApp through institutional contacts to staff in public-sector organisations that met the inclusion criteria. To protect institutional anonymity, participating organisations are not named; each was assigned an identification code instead.

Participants were invited to take part in a study of employees' experiences with AI decision support systems. The hypotheses and expected relationships between variables were not disclosed. The opening section of the questionnaire provided participant information, a definition of AI-DSS and an informed consent form. Respondents who did not give consent or did not meet the inclusion criteria were automatically screened out.

Participation was voluntary, and respondents were not asked to provide personally identifiable information, actual organisation names or confidential system names. This was intended to preserve anonymity and to allay concerns about disclosing organisational or system information. Because the survey link was distributed through institutional contacts and then forwarded by email and WhatsApp, the number of people who received it is unknown, and a conventional response rate could not be calculated.

4.11 Ethical considerations

The study adhered to established ethical principles throughout. Participation was voluntary and based on informed consent; respondents were informed of the study's purpose and of their right to withdraw at any time without consequence. No personal data were collected, and participants' privacy was protected throughout. The survey did not ask for organisation names, confidential system names, proprietary algorithms or any other sensitive information that could identify respondents. The data were used solely for academic research and were stored securely with restricted access.

4.12 Data screening and preparation

Data were screened according to predefined protocols before the main analyses. The initial dataset contained 461 responses. Following a review against the predefined data quality criteria, six cases that did not meet these criteria were removed, leaving a final analytic sample of 455. All retained respondents met the eligibility criteria and had given consent. There were no missing values, out-of-range values or duplicate records on either the substantive or the background variables.

Response quality was then assessed using several indicators, and no respondent was excluded on the basis of a single indicator. Completion times ranged from 285 to 1,180 seconds ($M = 534$ seconds), and no respondent completed the survey in less than half the mean time. One respondent selected the same response category for all 24 construct items, but their completion time (552 seconds) was close to the sample mean. Another respondent was identified as a multivariate outlier based on Mahalanobis D^2 ($p < .001$) but gave a varied response pattern. In addition, nine respondents reported an unusual combination of age and public-sector experience. Because no case was flagged by more than one indicator, all cases were retained for the main analysis; the ten flagged cases (the

uniform responder and the nine with unusual age-experience combinations) were examined in a robustness check (Section 5.9).

The distributions of the study variables were also examined. Skewness values ranged from -0.96 to -0.60 and kurtosis values from -0.67 to 0.10, within accepted thresholds for univariate normality ($|\text{skewness}| < 2$ and $|\text{kurtosis}| < 7$; Kline, 2016). However, Mardia's (1970) test indicated significant multivariate non-normality (multivariate skewness = 51.77, $p < .001$; multivariate kurtosis = 673.17, $z = 14.84$, $p < .001$). Robust estimation was therefore used in all subsequent analyses.

4.13 Strategy for data analysis

Analyses were conducted in R version 4.3.3 (R Core Team, 2024) using lavaan 0.6-17 (Rosseel, 2012) and semTools 0.5-6 (Jorgensen et al., 2022); descriptive statistics were computed with the psych package. Table 5 summarises the analytical stages and decision criteria.

4.13.1 Estimation

All CFA and SEM models were estimated using robust maximum likelihood (MLR), which provides Yuan-Bentler scaled test statistics and Huber-White robust standard errors under multivariate non-normality. Although seven-point indicators can be treated as approximately continuous, all models were re-estimated with the indicators treated as ordinal using WLSMV to check the stability of the results (Li, 2016).

4.13.2 Measurement model

A five-factor CFA was estimated in which each of the 24 items loaded on its intended factor and all factors were allowed to correlate. Model fit was assessed using the scaled χ^2 , robust CFI and TLI (ideally $\geq .95$), robust RMSEA with its 90% confidence interval (ideally $\leq .06$) and SRMR ($\leq .08$) (Hu & Bentler, 1999). Reliability was assessed using composite reliability (CR) and McDonald's ω ($\geq .70$), with Cronbach's α as a supplementary index. Convergent validity was assessed using standardised factor loadings (ideally $\geq .70$) and average variance extracted (AVE $\geq .50$) (Fornell & Larcker, 1981). Discriminant validity was assessed using the heterotrait-monotrait ratio (HTMT $< .85$; Henseler et al., 2015), the Fornell-Larcker criterion and 95% confidence intervals for the latent correlations. The five-factor model was also compared with alternative models in which conceptually adjacent constructs were merged, using the Satorra-Bentler scaled χ^2 difference test (Satorra & Bentler, 2001). Items were not dropped automatically for low loadings, and no residual correlations were added on the basis of modification indices alone (Gaskin et al., 2025).

4.13.3 Structural model and mediation

In the structural model, all seven direct paths were estimated simultaneously, with the three exogenous constructs allowed to correlate. For each path, unstandardised and standardised estimates, robust standard errors, z statistics, exact p values and 95% confidence intervals are reported, together with R^2 for DMQ and AE and Cohen's f^2 effect sizes (Cohen, 1988). The partial-mediation model was compared with a nested full-mediation model in which the three direct paths to AE were fixed to zero. Variance inflation factors (VIFs) were examined to rule out multicollinearity among the correlated predictors.

The three specific indirect effects (H8-H10) were estimated as the products of the corresponding path coefficients and tested using 10,000 non-parametric bootstrap resamples. An indirect effect was considered significant if its 95% percentile bootstrap confidence interval excluded zero (Preacher & Hayes, 2008). A significant direct effect was not required for mediation, and the type of mediation was classified following Zhao et al. (2010). Direct paths were compared using robust Wald tests, and indirect effects were compared using bootstrap confidence intervals of their differences.

4.13.4 Common method variance

Common method variance (CMV) was addressed using both procedural and statistical remedies (Podsakoff et al., 2003, 2012). Procedural remedies included respondent anonymity, omission of construct labels, separation of antecedent, mediator and outcome items into blocks, neutral wording, and assurances that there were no right or wrong answers. Statistically, Harman's single-factor test was conducted, the five-factor CFA was compared with a one-factor CFA, an unmeasured latent method construct (ULMC) was added to the model (Richardson et al., 2009), and full-collinearity VIFs were examined (Kock, 2015).

4.13.5 Robustness, non-response and clustering

The robustness of the findings was tested by re-estimating the model (a) with ordinal indicators using WLSMV, (b) with theoretically relevant control variables (frequency and duration of AI-DSS use and managerial position) as predictors of DMQ and AE, (c) excluding the ten flagged respondents (Section 4.12), and (d) with cluster-robust standard errors to account for the nesting of respondents within organisations. Age, gender and education were not included as controls because there was no theoretical reason to expect them to confound the hypothesised relationships. Non-response bias was assessed by comparing the first and last quartiles of respondents, ordered by submission time, on the five constructs and key background variables (Armstrong & Overton, 1977).

4.13.6 Measurement invariance and multigroup analysis

Configural, metric and scalar invariance were tested across managerial level (non-supervisory employees vs. supervisors and managers) and frequency of AI-DSS use (at least daily vs. less than daily). These two boundary conditions were identified from the literature and met the sample size requirements for multigroup confirmatory factor analysis. Invariance was evaluated using the scaled χ^2 difference test together with changes in CFI ($\Delta\text{CFI} \geq -.010$) and RMSEA ($\Delta\text{RMSEA} \leq .015$) (Cheung & Rensvold, 2002; Chen, 2007; Vandenberg & Lance, 2000).

Table 5. Analysis Stages and Decision Criteria

Stage	Analytical step	Decision criterion
1	Screening and data quality checks	Retain eligible respondents; no exclusion because of a single quality indicator
2	Descriptive statistics and normality	Skewness/kurtosis; Mardia's test → MLR
3	Five-factor CFA and model fit	CFI/TLI $\geq .95$; RMSEA $\leq .06$ (90% CI); SRMR $\leq .08$
4	Reliability and convergent validity	$\lambda \geq .70$ (ideal); $\alpha, \omega, CR \geq .70$; AVE $\geq .50$
5	Discriminant validity	HTMT $< .85$; $\sqrt{AVE} >$ inter-construct correlations; CI of correlations does not include 1
6	Alternative models	Merged-construct models have significantly poorer fit (scaled $\Delta\chi^2$)
7	Common method variance	Harman $< 50\%$; poor fit of one-factor CFA; ULMC; full-collinearity VIF < 3.3
8	Structural model	β , robust SE, 95% CI, R^2, f^2 ; partial vs. full mediation (scaled $\Delta\chi^2$)
9	Mediation	10,000 bootstrap resamples; 95% CI of indirect effect does not include 0
10	Robustness analyses	Stability with respect to WLSMV, controls, exclusions, and cluster-robust SE
11	Non-response bias	No differences between early and late responses
12	Measurement invariance	Non-significant $\Delta\chi^2$; $\Delta CFI \geq -.010$; $\Delta RMSEA \leq .015$

5. Findings

5.1 Response and data screening

As described in Section 4.12, the analytic sample comprised 455 eligible respondents from 22 organisations, after six of the 461 initial responses were removed. The sample exceeded the formal minimum of 341 by 114 cases and the planned sample of 400 by 55 cases. Because the dataset was complete, all analyses used the full sample.

5.2 Participant profile

Table 6 summarises the respondent profile. Respondents were predominantly male (275, 60.4%) and held a bachelor's (203, 44.6%) or master's degree (162, 35.6%). They spanned all job levels, including non-supervisory staff (193, 42.4%) and supervisors, middle managers and senior managers (248, 54.5%), and all seven organisation types, most commonly authorities and regulatory bodies (99, 21.8%) and ministries (95, 20.9%). The most frequently reported system types were AI-enabled business intelligence (96, 21.1%), generative AI (86, 18.9%) and predictive analytics (83, 18.2%). More than two-thirds of respondents (317, 69.7%) had used their focal system for at least six months, and more than one-third (168, 36.9%) used it at least daily.

Table 6. Respondent Characteristics (N = 455)

Characteristic	n	%
Gender		
Male	275	60.4
Female	171	37.6
Prefer not to say	9	2.0
Age group		
Under 30	40	8.8
30-39	151	33.2
40-49	148	32.5
50-59	81	17.8
60 or above	28	6.2
Prefer not to say	7	1.5
Highest qualification		

Characteristic	n	%
Diploma or less	23	5.1
Bachelor's	203	44.6
Master's	162	35.6
Doctorate	58	12.7
Other	9	2.0
Public-sector experience		
Less than 3 years	48	10.5
3-5 years	83	18.2
6-10 years	140	30.8
11-15 years	110	24.2
More than 15 years	74	16.3
Job level		
Professional/administrative staff (non-supervisory)	193	42.4
Supervisor/team leader	81	17.8
Middle management	118	25.9
Senior/executive management	49	10.8
Other	14	3.1
Organisation type		
Ministry	95	20.9
Authority/regulatory or independent body	99	21.8
Municipality	34	7.5
Public university/educational institution	78	17.1
Public health organisation	76	16.7
Judicial/security body	35	7.7
Other public organisation	38	8.4
Decision level supported		
Operational/day-to-day	154	33.8
Administrative/tactical	151	33.2
Strategic/policy	62	13.6
More than one level	88	19.3
Type of AI-DSS		
Predictive/forecasting analytics	83	18.2
Recommendation, prioritisation or risk assessment	70	15.4
AI-enabled business intelligence/analytics	96	21.1
Natural language processing	35	7.7
Generative AI for decision support	86	18.9
Combination of the above	70	15.4
Other AI-DSS	15	3.3
Duration of AI-DSS use		
Less than 3 months	59	13.0
3 to <6 months	79	17.4
6 months to <1 year	115	25.3
1 to <2 years	132	29.0
2 years or more	70	15.4

Characteristic	n	%
Frequency of AI-DSS use		
Less than once a month	63	13.8
1-3 times a month	88	19.3
Once to several times a week	136	29.9
About once a day	99	21.8
Several times a day	69	15.2
Self-rated AI experience		
Very basic	75	16.5
Basic	97	21.3
Intermediate	145	31.9
Advanced	96	21.1
Very advanced	42	9.2

Note. Percentages may not total 100% because of rounding. Because the sample was purposively selected, percentages describe the sample only.

5.3 Descriptive statistics and correlations

Construct means ranged from 5.20 to 5.39 on the seven-point scale, indicating generally favourable evaluations of the focal AI-DSS, and standard deviations of 1.31 to 1.42 reflected substantial individual variation (Table 7). DMQ received the highest rating ($M = 5.39$) and SQ the lowest ($M = 5.20$). All correlations between composite scores were positive and significant ($r = .33$ to $.57$, all $p < .001$), with the strongest between DMQ and AE and the weakest between SQ and TTF. These moderate correlations give no initial indication of construct redundancy.

At the item level, 66% to 75% of respondents gave a positive rating (categories 5-7) on each item, 25% to 35% selected the highest category, and fewer than 4% selected the lowest (Appendix Figure A1). These patterns show the negative skew expected among active users, without floor or ceiling effects severe enough to restrict variability, and they are similar across all five constructs.

Table 7. Means, Standard Deviations and Correlations of the Study Constructs

Construct	M	SD	Skew	Kurtosis	1	2	3	4	5
1. Information quality	5.25	1.42	-0.69	-0.40	(.90)				
2. System quality	5.20	1.31	-0.66	-0.24	.39***	(.86)			
3. Task-technology fit	5.28	1.41	-0.75	-0.20	.43***	.33***	(.88)		
4. Decision-making quality	5.39	1.33	-0.76	-0.08	.53***	.42***	.55***	(.85)	
5. Administrative efficiency	5.25	1.36	-0.67	-0.42	.49***	.39***	.52***	.57***	(.90)

Note. $N = 455$. Composite scores are item means (range 1-7). Cronbach's α is shown in parentheses on the diagonal. *** $p < .001$ (two-tailed).

5.4 Measurement model

The five-factor CFA fitted the data very well: $\chi^2(242) = 257.20$, $p = .240$, robust CFI = .997, TLI = .997, RMSEA = .012 [90% CI .000, .024] and SRMR = .031 (Table 8). All standardised loadings were significant ($p < .001$) and ranged from .667 to .855; only one item (SQ3, $\lambda = .667$) fell marginally below the ideal threshold of .70, and it was retained in line with the analysis plan (Section 4.13.2). Cronbach's α (.852 to .896), McDonald's ω (.854 to .897) and composite reliability (.853 to .897) all exceeded .70, and AVE exceeded .50 for every construct (.553 to .650), supporting reliability and convergent validity.

Table 8. Measurement Model: Item Statistics, Loadings, Reliability and Convergent Validity

Construct / item	M	SD	CITC	λ	α	ω	CR	AVE
Information quality					.896	.897	.897	.636
IQ1	5.19	1.73	.75	.805				
IQ2	5.28	1.68	.73	.787				
IQ3	5.30	1.65	.80	.855				
IQ4	5.25	1.67	.67	.709				
IQ5	5.24	1.71	.77	.823				

Construct / item	M	SD	CITC	λ	α	ω	CR	AVE
System quality					.860	.860	.861	.553
SQ1	5.18	1.66	.70	.763				
SQ2	5.31	1.57	.68	.738				
SQ3	5.22	1.65	.61	.667				
SQ4	5.24	1.60	.70	.780				
SQ5	5.06	1.69	.69	.764				
Task-technology fit					.881	.881	.881	.650
TTF1	5.29	1.60	.73	.789				
TTF2	5.35	1.64	.74	.813				
TTF3	5.22	1.68	.74	.802				
TTF4	5.25	1.64	.76	.821				
Decision-making quality					.852	.854	.853	.593
DMQ1	5.36	1.60	.69	.774				
DMQ2	5.44	1.57	.66	.736				
DMQ3	5.29	1.62	.67	.740				
DMQ4	5.45	1.61	.75	.826				
Administrative efficiency					.895	.896	.896	.591
AE1	5.27	1.74	.75	.799				
AE2	5.30	1.61	.77	.824				
AE3	5.13	1.73	.69	.741				
AE4	5.30	1.63	.67	.705				
AE5	5.30	1.61	.70	.747				
AE6	5.22	1.74	.74	.789				

Note. CITC = corrected item-total correlation; λ = standardised loading (MLR; all $p < .001$); ω = McDonald's omega; CR = composite reliability; AVE = average variance extracted. Model fit: $\chi^2(242) = 257.20$, $p = .240$, CFI = .997, TLI = .997, RMSEA = .012 [90% CI .000, .024], SRMR = .031.

Alternative measurement models that merged related constructs fitted considerably worse than the five-factor model (Table 9). Merging IQ and SQ ($\Delta\chi^2 = 789.80$, $\Delta df = 4$) and merging DMQ and AE ($\Delta\chi^2 = 330.90$, $\Delta df = 4$) both produced significantly poorer fit, confirming that the two IS success quality dimensions, and the mediator and outcome constructs, are empirically distinct.

Table 9. Comparison of Alternative Measurement Models

Model	χ^2	df	CFI	TLI	RMSEA	SRMR	$\Delta\chi^2$ (Δdf)	p
Five-factor (hypothesised)	257.20	242	.997	.997	.012	.031	-	-
Four-factor: IQ+SQ merged	917.33	246	.879	.864	.081	.079	789.80 (4)	< .001
Four-factor: DMQ+AE merged	630.46	246	.930	.922	.061	.055	330.90 (4)	< .001
Three-factor: IQ+SQ+TTF merged	1558.10	249	.765	.740	.112	.095	2116.81 (7)	< .001
One-factor	2270.89	252	.634	.599	.139	.108	1982.49 (10)	< .001

Note. All models estimated with MLR; χ^2 is the Yuan-Bentler scaled statistic, and CFI, TLI and RMSEA are robust versions. $\Delta\chi^2$ is the Satorra-Bentler scaled difference test against the five-factor model.

5.5 Discriminant validity

All three criteria supported discriminant validity (Table 10). The highest HTMT ratio was .654 (DMQ-AE), well below the conservative threshold of .85. The square root of each construct's AVE exceeded its correlations with all other constructs, satisfying the Fornell-Larcker criterion. The upper limit of the confidence interval for the highest latent correlation (DMQ-AE: .653, 95% CI [.578, .729]) was well below 1. The correlation between IQ and SQ was moderate (.447).

Appendix Figure A2 presents the same information graphically. Each diagonal element ($\sqrt{\text{AVE}}$) is darker than the other cells in its row and column, and the HTMT ratios in the upper triangle closely mirror the latent correlations in the lower triangle.

Table 10. Discriminant Validity: Fornell-Larcker Criterion and HTMT Ratios

Construct	IQ	SQ	TTF	DMQ	AE
Information quality	.797	.436	.473	.604	.546
System quality	.447	.744	.374	.491	.443
Task-technology fit	.485	.380	.806	.636	.578
Decision-making quality	.600	.494	.634	.770	.654
Administrative efficiency	.545	.446	.582	.653	.768

Note. Bold diagonal values are square roots of AVE. Values below the diagonal are latent (CFA) correlations; italicised values above the diagonal are HTMT ratios.

5.6 Common method variance

The statistical tests consistently indicated that common method variance is not a serious concern. A single factor accounted for 36.8% of the variance in the 24 items (principal components analysis; 39.4% with principal axis factoring), below the 50% threshold. More importantly, the one-factor CFA fitted the data poorly (CFI = .634, RMSEA = .139, SRMR = .108) compared with the five-factor model (Table 9). When a ULMC with equal loadings was added to the model, the estimated method variance was negative, an inadmissible (Heywood) solution, and model fit improved only trivially ($\Delta\text{CFI} = .001$); with freely estimated loadings, the model failed to converge. Both outcomes suggest that there is no identifiable common factor beyond the substantive constructs, and the parameters of the inadmissible solution are not interpreted. Full-collinearity VIFs ranged from 1.31 to 1.93, below the 3.3 threshold (Kock, 2015), and no correlation approached .90. Nevertheless, because these tests are post hoc, method variance cannot be ruled out entirely (Podsakoff et al., 2012).

5.7 Structural model and hypothesis testing

Because the partial-mediation structural model is saturated in its latent relationships, its fit is identical to that of the CFA: $\chi^2(242) = 257.20$, $p = .240$, CFI = .997, TLI = .997, RMSEA = .012 [90% CI .000, .024], SRMR = .031. The full-mediation model, in which the paths from IQ, SQ and TTF to AE were constrained to zero, fitted significantly worse (scaled $\Delta\chi^2(3) = 33.41$, $p < .001$; robust CFI = .992; robust RMSEA = .021; AIC = 35,997.8 vs. 35,965.8), so the partial-mediation model was retained. The model explained 54.5% of the variance in DMQ and 50.7% of the variance in AE, both substantial proportions. Multicollinearity was not a concern, as all VIFs were moderate (≤ 1.93).

All three AI-DSS characteristics were positively and significantly associated with DMQ (Table 11 and Figure 3). The association was strongest for TTF ($\beta = .407$, $p < .001$, $f^2 = .221$), followed by IQ ($\beta = .313$, $p < .001$, $f^2 = .097$) and SQ ($\beta = .200$, $p < .001$, $f^2 = .047$). H1, H2 and H3 were therefore supported. Robust Wald tests showed that the TTF $\rightarrow$ DMQ path was significantly stronger than the SQ $\rightarrow$ DMQ path ($\chi^2(1) = 5.59$, $p = .018$), whereas the differences between TTF and IQ ($\chi^2(1) = 0.98$, $p = .323$) and between IQ and SQ ($\chi^2(1) = 1.74$, $p = .187$) were not significant.

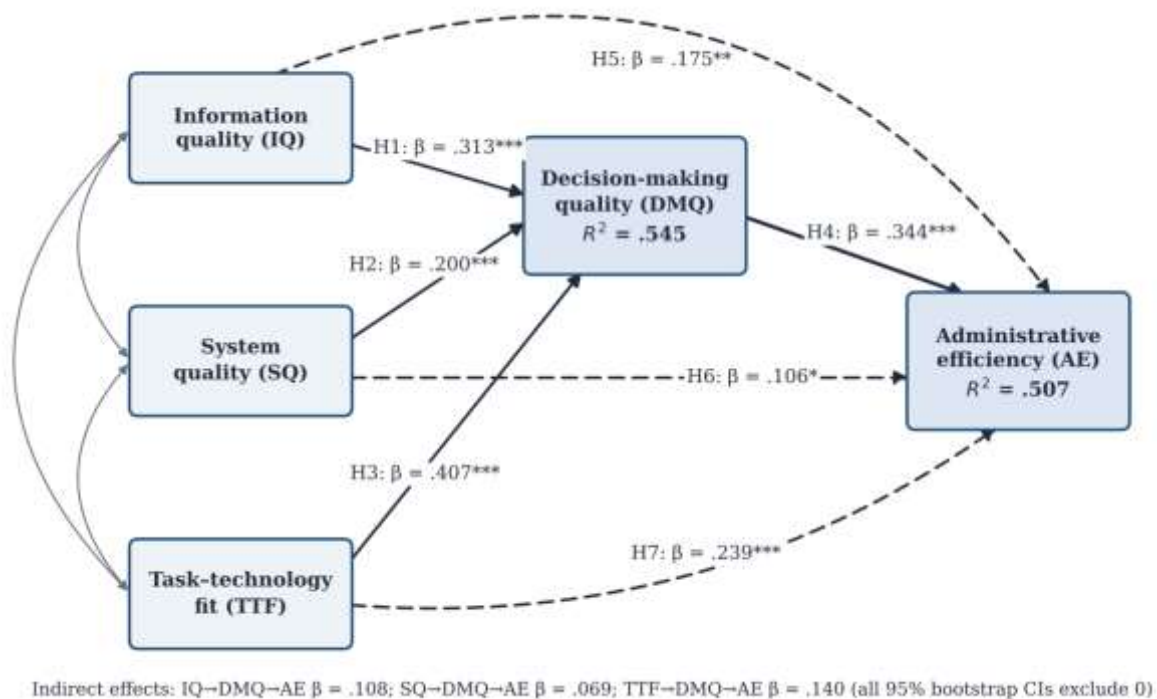
DMQ was the strongest predictor of AE ($\beta = .344$, $p < .001$, $f^2 = .109$), supporting H4. The direct effects of the AI-DSS characteristics on AE were smaller. TTF ($\beta = .239$, $p < .001$, $f^2 = .030$) and IQ ($\beta = .175$, $p = .002$, $f^2 = .014$) had significant direct effects on AE, supporting H7 and H5, respectively. The direct effect of SQ on AE was weak ($\beta = .106$, $p = .036$, $f^2 = .007$); H6 was therefore supported, but only weakly, and Section 5.9 shows that this path was not robust.

Table 11. Structural Model Results and Hypothesis Testing (H1-H7)

Hypothesis	Path	B	SE	β	95% CI (β)	z	p	f^2	Decision
H1	IQ $\rightarrow$ DMQ	0.280	0.047	.313	[.209, .417]	5.91	< .001	.097	Supported
H2	SQ $\rightarrow$ DMQ	0.196	0.054	.200	[.102, .298]	4.00	< .001	.047	Supported
H3	TTF $\rightarrow$ DMQ	0.400	0.060	.407	[.301, .512]	7.54	< .001	.221	Supported
H4	DMQ $\rightarrow$ AE	0.384	0.093	.344	[.193, .495]	4.46	< .001	.109	Supported
H5	IQ $\rightarrow$ AE	0.175	0.057	.175	[.066, .285]	3.15	.002	.014	Supported
H6	SQ $\rightarrow$ AE	0.117	0.057	.106	[.007, .206]	2.09	.036	.007	Supported (weak; not robust)
H7	TTF $\rightarrow$ AE	0.263	0.070	.239	[.118, .360]	3.87	< .001	.030	Supported

Note. B = unstandardised coefficient; SE = robust standard error (MLR); β = standardised coefficient; f^2 = Cohen's effect size (.02 small, .15 medium, .35 large). $R^2(\text{DMQ}) = .545$; $R^2(\text{AE}) = .507$. Model fit: $\chi^2(242) = 257.20$, $p = .240$, CFI = .997, TLI = .997, RMSEA = .012 [90% CI .000, .024], SRMR = .031.

Figure 3. Structural Model with Standardised Path Coefficients



Note. Standardised MLR estimates. Solid arrows denote the decision-making quality pathway, and dashed arrows denote direct paths. * $p < .05$, ** $p < .01$, *** $p < .001$. Measurement indicators and exogenous covariances are omitted for clarity.

5.8 Mediation analysis

The bootstrap results (10,000 resamples) supported all three mediation hypotheses (Table 12). The indirect effect of TTF on AE through DMQ was the strongest ($\beta = .140$, 95% CI [.070, .221]), followed by those of IQ ($\beta = .108$, 95% CI [.054, .173]) and SQ ($\beta = .069$, 95% CI [.029, .120]). No confidence interval included zero, so H8, H9 and H10 were supported. Because each characteristic also retained a positive direct effect on AE, the mediation was complementary (partial) in each case (Zhao et al., 2010). DMQ accounted for 37% to 39% of the total effect of each characteristic on AE. Together, the three indirect effects summed to $\beta = .316$, smaller than the sum of the direct effects ($\beta = .521$); neither confidence interval included zero. Bootstrap comparisons showed that the indirect effect through TTF was significantly stronger than that through SQ ($\Delta\beta = .071$; the 95% bootstrap CI of the difference excluded zero), whereas the TTF-IQ and IQ-SQ differences were not significant.

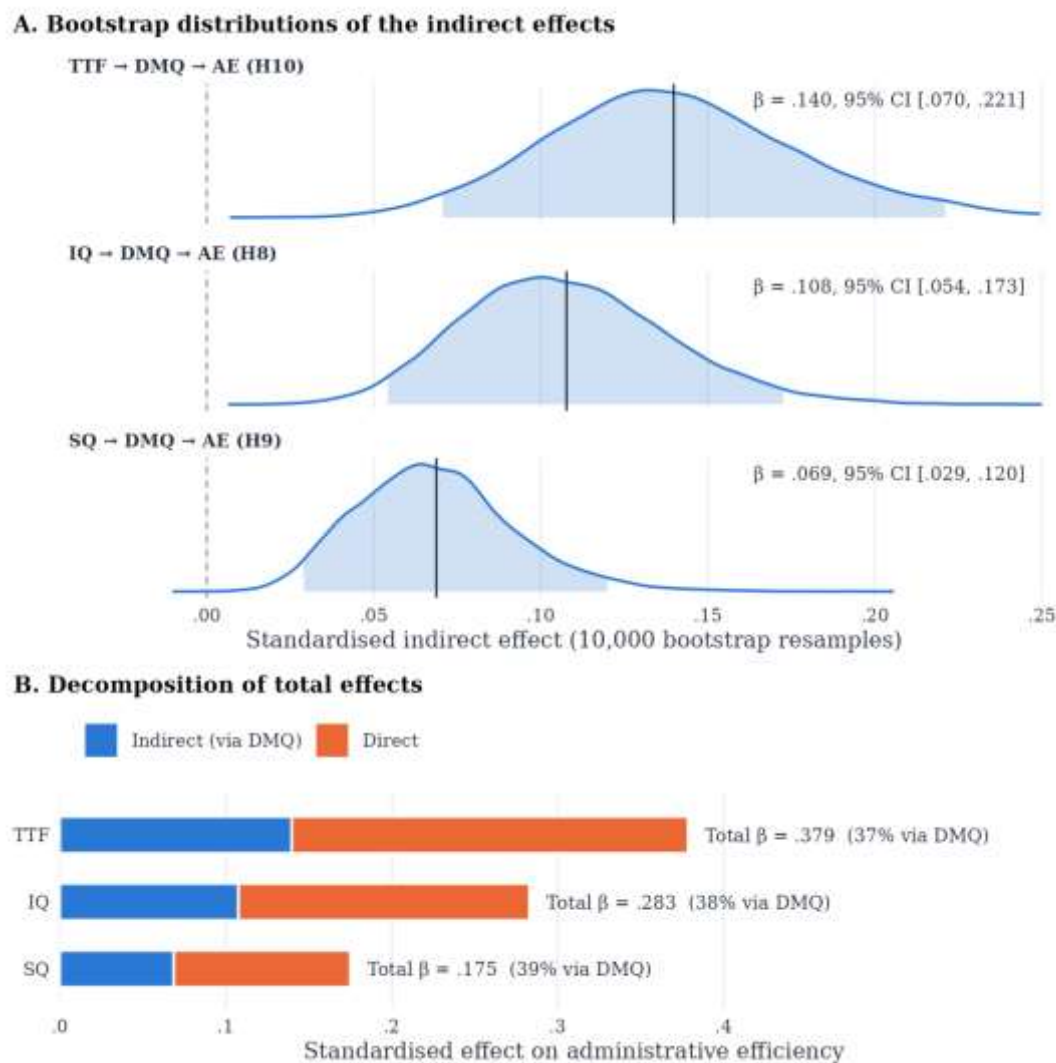
Table 12. Direct, Indirect and Total Effects: Bootstrap Mediation Results (H8-H10)

Hypothesis	Path	Indirect β	Boot SE	95% CI (indirect β)	Indirect B [95% CI]	Direct β	Total β	VAF	Mediation type
H8	IQ → DMQ → AE	.108	.031	[.054, .173]	.107 [.053, .176]	.175**	.283***	38%	Complementary (partial)
H9	SQ → DMQ → AE	.069	.024	[.029, .120]	.075 [.031, .136]	.106*	.175***	39%	Complementary (partial)
H10	TTF → DMQ → AE	.140	.038	[.070, .221]	.154 [.077, .246]	.239***	.379***	37%	Complementary (partial)

Note. Percentile bootstrap confidence intervals based on 10,000 resamples (ML estimates); bias-corrected intervals yielded similar results. VAF = variance accounted for (indirect effect/total effect). * $p < .05$, ** $p < .01$, *** $p < .001$.

Figure 4 illustrates these findings. Panel A shows the bootstrap sampling distributions of the three standardised indirect effects. The distributions are clearly right-skewed, and the distribution for TTF lies furthest from zero. Panel B decomposes the total effect of each characteristic on AE into its direct component and its indirect component through DMQ, showing that the share carried by DMQ is nearly identical across the three characteristics (37% to 39%).

Figure 4. Mediation by Decision-Making Quality: Bootstrap Distributions and Effect Decomposition



Note. Panel A shows the bootstrap distribution of each standardised indirect effect (10,000 resamples); the shaded region marks the 95% percentile confidence interval, the solid line the point estimate and the dashed line zero. Panel B shows the standardised direct and indirect components of each total effect (MLR estimates); percentages give the VAF, the share of the total effect transmitted through DMQ.

5.9 Robustness and non-response analyses

The substantive conclusions were robust across alternative specifications (Table 13). Treating the indicators as ordinal (WLSMV; CFI = .998, RMSEA = .013), adding frequency of use, duration of use and managerial status as controls, excluding the ten flagged respondents, and applying cluster-robust standard errors for the 22 organisations all left H1-H5 and H7-H10 significant, with coefficients close to the baseline estimates. H6 (the direct SQ → AE path), however, was fragile: it became non-significant when the controls were added ($\beta = .092$) and when cluster-robust standard errors were applied. Clustering effects were negligible, with intraclass correlations of construct scores within organisations close to zero ($ICC \leq .011$). Among the controls, only duration of use significantly predicted AE, and the association was negative ($\beta = -.123$, $p = .001$); frequency of use ($\beta = .074$, $p = .060$) and managerial status ($\beta = .024$, $p = .533$) were not significant.

Table 13. Robustness of Standardised Estimates Across Alternative Specifications

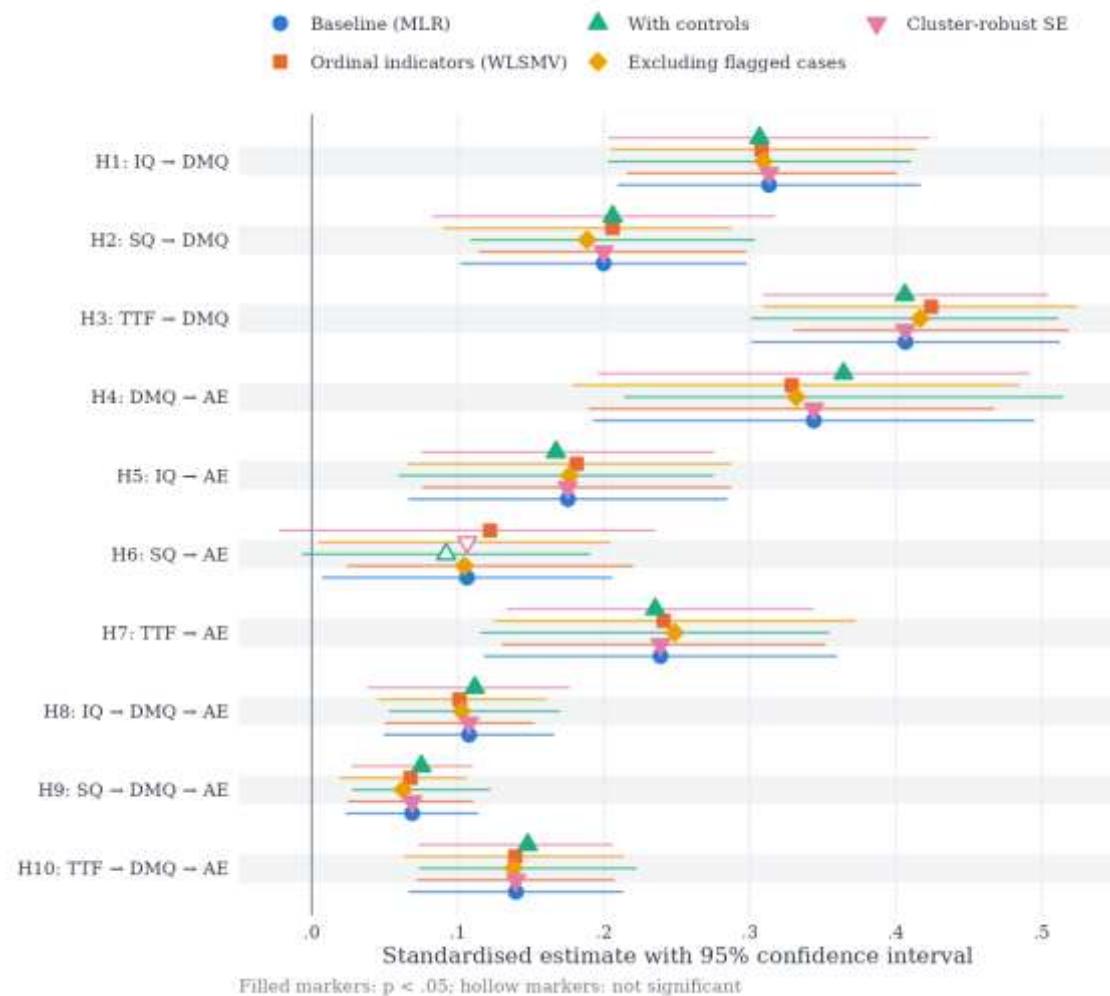
Specification	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10
Baseline (MLR, N = 455)	.313***	.200***	.407***	.344***	.175**	.106*	.239***	.108***	.069**	.140***
Ordinal indicators (WLSMV)	.308***	.206***	.424***	.329***	.181***	.122*	.241***	.101***	.068**	.139***
With controls	.307***	.206***	.406***	.364***	.167**	.092	.235***	.112***	.075**	.148***

Specification	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10
Excluding flagged cases (N = 445)	.309***	.189***	.417***	.332***	.177**	.105*	.249***	.103***	.063**	.138***
Cluster-robust SE	.313***	.200***	.407***	.344***	.175***	.106	.239***	.108**	.069**	.140***

Note. Standardised estimates. H1-H7 are direct paths; H8-H10 are indirect effects. Controls: AI-DSS use frequency, duration of use and managerial status. Flagged cases: one uniform responder and nine respondents with an unusual age-experience combination. * $p < .05$, ** $p < .01$, *** $p < .001$.

Figure 5 plots the point estimates and 95% confidence intervals across the five specifications. For every hypothesis except H6, the estimates cluster closely and all confidence intervals exclude zero. H6 (SQ → AE) is the only path whose confidence interval crosses zero, which it does in two specifications, shown by the hollow markers.

Figure 5. Robustness of Hypothesised Paths and Indirect Effects Across Five Model Specifications



Note. Points are standardised estimates and horizontal lines are 95% confidence intervals (delta-method intervals from each specification). Filled markers indicate $p < .05$; hollow markers indicate non-significant estimates. Marker shapes distinguish the specifications, so the figure can be read without colour; Table 13 reports the same values.

The first and last quartiles of respondents by submission time ($n = 113$ each) did not differ significantly on any construct (all $p \geq .20$; Table 14) or on job level, organisation type, gender or frequency of use (all χ^2 tests, $p > .30$). There is therefore little evidence of non-response bias, although this comparison is not a definitive test (Armstrong & Overton, 1977).

Table 14. Comparison of Early and Late Respondents

Construct	Early M	Late M	t	df	p
Information quality	5.12	5.18	-0.28	222.8	.780
System quality	5.26	5.05	1.29	222.8	.200
Task-technology fit	5.16	5.30	-0.74	223.3	.462

Construct	Early M	Late M	t	df	p
Decision-making quality	5.35	5.27	0.39	224.0	.698
Administrative efficiency	5.12	5.25	-0.70	207.8	.484

Note. Early and late groups are the first and last quartiles of respondents by submission time; Welch's t-tests.

5.10 Measurement invariance and multigroup comparison

Measurement invariance was tested across managerial level (non-supervisory staff, $n = 193$, vs. supervisors and managers, $n = 248$; the 14 respondents reporting other job levels were excluded from this comparison) and frequency of AI-DSS use (daily or more, $n = 168$, vs. less than daily, $n = 287$). In both comparisons, the configural model fitted well. Constraining factor loadings (metric invariance) and then item intercepts (scalar invariance) produced non-significant scaled χ^2 differences, and changes in CFI and RMSEA were well within the .010 and .015 cut-offs (Table 15). Full scalar invariance was therefore established, indicating that the constructs were measured equivalently across groups. Constraining all seven structural paths to equality did not significantly worsen fit for either managerial level ($\Delta\chi^2(7) = 3.42$, $p = .844$) or frequency of use ($\Delta\chi^2(7) = 6.60$, $p = .472$).

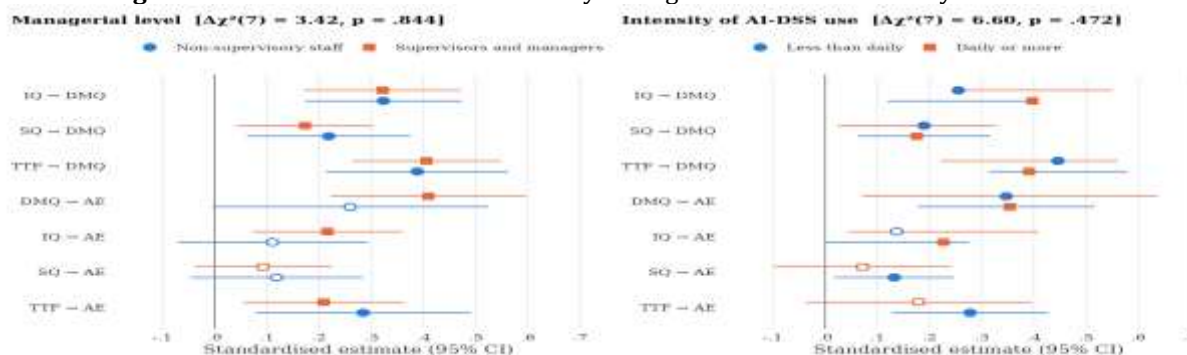
Table 15. Measurement Invariance Across Managerial Level and Intensity of AI-DSS Use

Group / model	χ^2	df	CFI	RMSEA	SRMR	$\Delta\chi^2$ (Δ df)	p	Δ CFI
Managerial level								
Configural	536.89	484	.990	.023	.042	-	-	-
Metric	552.84	503	.991	.022	.044	15.29 (19)	.704	.001
Scalar	574.85	522	.990	.022	.045	22.09 (19)	.280	-.001
Intensity of AI-DSS use								
Configural	518.04	484	.994	.018	.040	-	-	-
Metric	536.03	503	.994	.018	.044	17.66 (19)	.545	.000
Scalar	548.71	522	.995	.016	.044	12.20 (19)	.877	.001

Note. MLR estimation; scaled χ^2 and robust CFI/RMSEA. $\Delta\chi^2$ is the Satorra-Bentler scaled difference test against the preceding, less constrained model.

Figure 6 shows the path estimates for each group, with factor loadings constrained to be equal. The confidence intervals overlap substantially for every path, consistent with the non-significant tests of path equality. Some paths are significant in one group but not the other; for example, the DMQ $\rightarrow$ AE path narrowly misses significance for non-supervisory staff ($\beta = .259$, $p = .055$) but is .409 for supervisors and managers. Because the equality tests were non-significant, such differences are most plausibly attributable to the smaller subgroup samples.

Figure 6. Standardised Structural Paths by Managerial Level and Intensity of AI-DSS Use



Note. Multigroup SEM with factor loadings constrained equal (metric invariance). Points are standardised estimates and lines 95% confidence intervals; hollow markers indicate $p \geq .05$. Panel headings report the scaled χ^2 difference test for constraining all seven paths to equality.

5.11 Summary of hypothesis tests

Table 16. Summary of Hypothesis Testing

Hypothesis	Relationship	Evidence	Result
H1	IQ $\rightarrow$ DMQ	$\beta = .313$, $p < .001$	Supported
H2	SQ $\rightarrow$ DMQ	$\beta = .200$, $p < .001$	Supported
H3	TTF $\rightarrow$ DMQ	$\beta = .407$, $p < .001$	Supported
H4	DMQ $\rightarrow$ AE	$\beta = .344$, $p < .001$	Supported

Hypothesis	Relationship	Evidence	Result
H5	IQ → AE	$\beta = .175, p = .002$	Supported
H6	SQ → AE	$\beta = .106, p = .036$	Supported (weak; not robust to controls or clustering)
H7	TTF → AE	$\beta = .239, p < .001$	Supported
H8	IQ → DMQ → AE	Indirect $\beta = .108, 95\% \text{ CI } [.054, .173]$	Supported (complementary partial mediation)
H9	SQ → DMQ → AE	Indirect $\beta = .069, 95\% \text{ CI } [.029, .120]$	Supported (complementary partial mediation)
H10	TTF → DMQ → AE	Indirect $\beta = .140, 95\% \text{ CI } [.070, .221]$	Supported (complementary partial mediation)

6. Discussion

This study examined whether, and how, the information quality, system quality and task-technology fit of AI-based DSS relate to administrative efficiency in Saudi public-sector organisations. The results support a decision-centred explanation of AI-DSS value. All three characteristics were associated with decision-making quality, which in turn was the strongest predictor of efficiency, and decision-making quality mediated the effect of each characteristic on efficiency. The mediation was partial: task-technology fit and information quality also had robust direct associations with efficiency, whereas the direct association for system quality was weak and not robust. This section discusses the results in relation to each research question.

6.1 AI-DSS characteristics and decision-making quality (RQ1)

The first research question concerned the relationships between the information quality, system quality and TTF of AI-DSS and DMQ. All three constructs were positively related to DMQ and together explained more than half (54.5%) of its variance. This result is consistent with the IS success model, in which the quality of a system and its outputs is the starting point for individual impact (DeLone & McLean, 1992, 2003). It also supports prior evidence linking analytics and decision support capabilities to decision-making effectiveness (Popović et al., 2012; Cao et al., 2015; Ghasemaghaei et al., 2018) and extends that evidence to AI-DSS in public administration.

The most notable finding is that task-technology fit was the strongest correlate of decision quality ($\beta = .407$), significantly stronger than system quality. This is what TTF theory predicts: technology improves performance not because of its sophistication per se but because its capabilities match task requirements (Goodhue & Thompson, 1995; Ziguers & Buckland, 1998). The point is especially pertinent for AI, because AI suits some types of decisions better than others (Jarrahi, 2018; Shrestha et al., 2019), and evidence that additional capability is not always beneficial (Aljukhadar et al., 2014) suggests that technologically advanced AI-DSS may fail to improve decisions when poorly matched to the task. How well a system supports the specific decision task was thus at least as important as the quality of its outputs.

Information quality was the second strongest antecedent ($\beta = .313$). Its link with decision quality supports the long-standing argument that good information leads to good decisions when decision-makers are able to use it (Raghunathan, 1999). The finding is particularly relevant in the AI context. Research on algorithmic advice in government shows that decision-makers may use advice selectively, accepting suggestions that confirm their own judgement and disregarding others (Alon-Barkat & Busuioc, 2023; Selten et al., 2023). Good information is therefore necessary but not sufficient for decision quality: it must also be trusted and incorporated into judgement. Because information quality was measured here in terms of the accuracy, relevance and clarity of outputs, the finding suggests that the interpretability of AI outputs is part of what users experience as information quality.

System quality had the weakest, though still significant, relationship with DMQ ($\beta = .200$). One explanation is that availability, responsiveness and reliability act as enablers rather than direct sources of judgement quality: they reduce the effort of using the system, and decision-makers tend to adopt the strategies that their tools make least effortful (Todd & Benbasat, 1999). This is consistent with syntheses of IS success research showing that system quality receives weaker support than information quality as a predictor of net benefits (Petter et al., 2008; Jeyaraj, 2020).

6.2 Decision-making quality and administrative efficiency (RQ2)

The second research question examined the relationship between decision quality and administrative efficiency. DMQ was the strongest predictor of AE ($\beta = .344$) after controlling for the direct effects of the three AI-DSS characteristics. This confirms the expectation that good decisions precede reduced delay, error and rework, and it is consistent with research that treats decision effectiveness as the intermediate process through which IT creates value (Melville et al., 2004; Cao et al., 2015; Ghasemaghaei et al., 2018). In public administration, where efficiency is only one of several performance dimensions (Boyne, 2003; Andersen et al., 2016), the result implies that efficiency gains from AI-DSS derive largely from improving the judgements that underpin administration, rather than from process automation alone. Because both constructs are self-reported, however, this relationship indicates a coherent perceived mechanism rather than objectively measured productivity gains.

6.3 Direct associations with administrative efficiency (RQ3)

The third research question examined whether AI-DSS characteristics are related to efficiency beyond their relationship with decision quality. Task-technology fit ($\beta = .239$) and information quality ($\beta = .175$) had significant direct relationships with AE. The direct relationship for system quality ($\beta = .106$) was weak and became non-significant when frequency of use, duration of use and managerial status were controlled or when cluster-robust standard errors were applied. The full-mediation model fitted significantly worse than the partial-mediation model, indicating that decision quality is an important channel but not the only one.

The direct TTF path implies that fit operates in two ways: it improves decisions by enabling better judgements, and it streamlines the execution of decision-related tasks by reducing searching, re-entry and checking, regardless of the quality of the resulting decision. This dual role is consistent with the original link between fit and individual performance in TTF theory (Goodhue & Thompson, 1995). Similarly, the direct IQ path implies that timely and accurate outputs save time and prevent unnecessary corrections even when the decision itself would be the same. For system quality, by contrast, the findings suggest a largely indirect effect: once its contribution to decision quality is accounted for, system quality adds little to efficiency. This is consistent with the view that object-based system qualities influence outcomes through users' subsequent beliefs and behaviours (Wixom & Todd, 2005). The weak SQ $\rightarrow$ AE relationship is therefore informative rather than a failed test: technical quality is a precondition for, but not a source of, efficiency gains from AI-DSS.

6.4 The mediating role of decision-making quality (RQ4)

The fourth research question asked whether decision-making quality mediates the relationship between AI-DSS characteristics and administrative efficiency. All three indirect effects were significant, and DMQ accounted for 37-39% of the total relationship between each characteristic and efficiency. The mediation was complementary, as the indirect and direct effects had the same sign (Zhao et al., 2010). Decision quality is thus a primary, but not the only, mediator. These findings respond to calls to open the "black box" linking IT and performance and specify a concrete step between individual-level and organisational-level impact that the IS success model leaves unexplained (DeLone & McLean, 2003; Melville et al., 2004; Gable et al., 2008).

The indirect effect through TTF was the strongest and was significantly greater than that through system quality. Taken together with the findings for RQ1 and RQ3, this suggests two distinct pathways. In the judgement pathway, information quality and TTF improve the quality of decisions, which in turn improves efficiency. In the execution pathway, information quality and TTF make task performance faster and less error-prone directly. System quality contributes mainly through the judgement pathway, acting as a facilitating condition.

6.5 Boundary conditions and the Saudi public-sector context

The measurement model was invariant, and the structural relationships did not differ significantly between non-managerial and managerial staff or between frequent and infrequent users (Figure 6). The mechanism therefore appears to operate similarly across organisational levels and intensities of use, which is noteworthy because managers and non-managers hold different decision authority and might be expected to respond differently to algorithmic advice (Shrestha et al., 2019; Alon-Barkat & Busuic, 2023). An unexpected finding from the robustness analysis was that longer use of the focal system was associated with slightly lower perceived efficiency ($\beta = -.123$) after controlling for AI-DSS characteristics. This may reflect initial efficiency gains becoming the new baseline, longer-used systems belonging to older technology generations, or accumulated experience changing how users interact with algorithms (Dietvorst et al., 2015; Logg et al., 2019).

The Saudi context makes these findings particularly relevant. AI adoption in the Saudi government has been strongly policy-driven, and prior research has focused largely on how AI is integrated into public organisations and on the role of leadership in that process (Alshahrani et al., 2022, 2025). Where AI implementation is mandated from the top down, there is a risk that implementation is equated with performance. This study shows instead that, for actual users, the value of an AI-DSS depends on its fit with decision tasks and the quality of its outputs, and that this value is realised largely through better decisions.

7. Theoretical Contributions

The study makes four theoretical contributions, each grounded in the empirical findings.

First, the study shifts research on public-sector AI from adoption to the post-adoption stage. Much of the literature on AI in government has addressed readiness, capabilities and the antecedents of adoption (Mikalef et al., 2022; Madan & Ashok, 2023), whereas evidence on the consequences of use remains scarce (Zuiderwijk et al., 2021). The study responds to calls to examine the post-adoption stage, where value is created or destroyed (Jasperson et al., 2005; Burton-Jones & Straub, 2006).

Second, the study integrates the IS success model and TTF theory around an outcome rather than around use. Previous integrations of the two theories have mostly sought to explain utilisation (Dishaw & Strong, 1999). The present model shows that artefact-centred and fit-centred theories offer complementary explanations of value: the IS success characteristics and task-technology fit each made a distinct contribution to decision quality, with fit showing the strongest association. Neither theory on its own could account for this pattern.

Third, the study specifies and empirically demonstrates decision quality as a mechanism through which AI-DSS characteristics influence administrative efficiency. The IS success model treats individual and organisational impacts as broad categories without detailing how they are connected (DeLone & McLean, 2003; Petter et al.,

2008). By positioning decision quality between system characteristics and administrative efficiency, and by showing that it carries more than a third of the effect of each characteristic on efficiency, the study clarifies this relationship for decision support systems.

Fourth, the study extends evidence on AI-DSS to the Saudi public sector in a theoretically informed way. The Saudi setting matters not merely as a location but because policy-driven adoption makes the gap between use and effectiveness critical. Showing that the same mechanism holds across managerial levels and intensities of use, with full measurement invariance, supports a context-sensitive account of AI-DSS value realisation in public administration that is robust across these subgroups.

8. Practical and Managerial Implications

Several recommendations follow for those who select, implement and manage AI-DSS in public organisations. Each is based on associations observed in this study and should be read in light of its cross-sectional design.

Government executives and digital transformation leaders should evaluate AI-DSS investments in terms of decision quality, not only deployment or usage statistics. Decision quality was the strongest correlate of efficiency and carried a substantial share of the value associated with each system characteristic, so measures of the consistency, accuracy and reliability of AI-supported decisions should be included in evaluations alongside cost and time measures.

Task-technology fit should be a primary criterion in AI-DSS selection. Because fit was strongly associated with both decision quality and efficiency, procurement should begin with a careful analysis of the decision tasks the system is intended to support, including their nature and information requirements, rather than with a comparison of product features. Pilot evaluations in which prospective users perform real decision tasks with the system offer one way to assess fit.

System developers and data analysts should treat information quality as a core design criterion. In this study, information quality encompassed the accuracy, timeliness, relevance and clarity of outputs, and it was associated with both decision quality and efficiency. Investment in data quality, decision-specific outputs and interpretable predictions and recommendations is therefore likely to yield good returns.

IT operations teams should maintain system availability, responsiveness and reliability, while recognising that these factors contribute to efficiency mainly by enabling better decisions. However well a system performs technically, efficiency gains are unlikely if its outputs are poor or its functions are poorly aligned with the work.

Public-sector decision-makers and their managers should deliberately integrate AI-DSS outputs into decision processes. Because efficiency gains are realised through decisions, training should focus on how to interpret, evaluate and combine system outputs with professional judgement.

9. Policy Implications

At the policy level, the findings support a shift from deployment-oriented to value-oriented evaluation of government AI systems. Counts of AI-DSS deployed and of employees trained to use them are input measures, not outcomes. Because efficiency gains from AI-DSS depend on fit and information quality and arise largely through better decisions, frameworks for evaluating government AI investments should include measures of decision quality and efficiency. Procurement processes could also make task-fit analysis and explicit information quality criteria, including data provenance, timeliness and interpretability of outputs, prerequisites for purchasing an AI-DSS.

The findings also have implications for responsible human involvement in decision-making. The study examined AI as decision support, in which a human makes the decision on the basis of system outputs; value therefore depends on the quality of human judgement. Human-in-the-loop policies that ensure decision-makers receive intelligible AI-DSS outputs are thus consistent with efficiency goals. These conclusions rest on perceptual, cross-sectional data and should be validated using objective performance measures.

10. Limitations and Future Research

This study has several limitations. First, the cross-sectional design allows the hypothesised pathways to be tested for consistency with the data but cannot establish causality; reverse pathways (for example, more efficient users forming more favourable perceptions of the system) cannot be ruled out. Second, administrative efficiency was measured as perceived efficiency without objective indicators, so the results describe how users perceive the benefits of AI-DSS. Third, all variables were self-reported by the same respondents. Procedural remedies were applied and several statistical tests showed no evidence of a dominant method factor, but method variance cannot be ruled out entirely. Fourth, purposive sampling and recruitment through institutional contacts by email and WhatsApp mean that the sample may over-represent engaged users, and the response rate could not be determined. Fifth, respondents evaluated different types of AI-DSS, from predictive analytics to generative AI, and the mechanism may vary by system type. Sixth, the study was conducted in a single country, and the policy-driven nature of AI implementation in Saudi Arabia may limit generalisability to other administrative systems. Seventh, although clustering of respondents within organisations was negligible, the design did not permit multilevel modelling of organisational factors. Finally, some items were adapted or developed for this study; although they performed well psychometrically, their validity requires further confirmation in other settings.

These limitations suggest several directions for future research. Longitudinal or panel designs could establish temporal precedence among quality, fit, decision quality and efficiency. Objective indicators such as processing time, error counts, system usage logs and supervisor ratings would allow perceived and actual efficiency to be compared. Matched multi-source designs or field experiments that vary the quality and fit of AI outputs would provide stronger tests of causality. Comparative studies across system types, sectors and countries could test the generalisability of the judgement and execution pathways. Where theoretically justified, future work could also examine the moderating roles of explainability, trust in AI and decision discretion, and investigate the unexpected negative association between duration of use and efficiency.

11. Conclusion

This study examined whether and how the characteristics of AI-based decision support systems relate to administrative efficiency in Saudi public-sector organisations. Integrating the IS success model and TTF theory, it proposed that decision quality mediates the effects of information quality, system quality and task-technology fit on efficiency. Covariance-based structural equation modelling of survey responses from 455 public-sector employees and managers who use AI-DSS supported this decision-centred account. Information quality, system quality and task-technology fit were all positively associated with decision quality, with task-technology fit showing the strongest association. Decision quality was the strongest predictor of efficiency and partially mediated the relationships between all three characteristics and efficiency.

These findings help shift research on public-sector AI from adoption to value realisation and show that quality-based and fit-based accounts of information system value are complementary. The main lesson for AI-enabled public administration is that implementing AI is not enough. Value depends on selecting systems that fit the decisions officials make and that produce high-quality outputs, and on recognising that efficiency gains flow largely from better decisions.

References

1. Ahn, M. J., & Chen, Y.-C. (2022). Digital transformation toward AI-augmented public administration: The perception of government employees and the willingness to use AI in government. *Government Information Quarterly*, 39(2), 101664. <https://doi.org/10.1016/j.giq.2021.101664>
2. Aljukhadar, M., Senecal, S., & Nantel, J. (2014). Is more always better? Investigating the task-technology fit theory in an online user context. *Information & Management*, 51(4), 391-397. <https://doi.org/10.1016/j.im.2013.10.003>
3. Alon-Barkat, S., & Busuioc, M. (2023). Human-AI interactions in public sector decision making: "Automation bias" and "selective adherence" to algorithmic advice. *Journal of Public Administration Research and Theory*, 33(1), 153-169. <https://doi.org/10.1093/jopart/muac007>
4. Alshahrani, A., Dennehy, D., & Mäntymäki, M. (2022). An attention-based view of AI assimilation in public sector organizations: The case of Saudi Arabia. *Government Information Quarterly*, 39(4), 101617. <https://doi.org/10.1016/j.giq.2021.101617>
5. Alshahrani, A., Griva, A., Dennehy, D., & Mäntymäki, M. (2024). Artificial intelligence and decision-making in government functions: Opportunities, challenges and future research. *Transforming Government: People, Process and Policy*, 18(4), 678-698. <https://doi.org/10.1108/TG-06-2024-0131>
6. Alshahrani, A., Griva, A., Dennehy, D., & Mäntymäki, M. (2025). The role of leadership and communication in AI assimilation: Case studies from Saudi Arabia's public sector organizations. *Transforming Government: People, Process and Policy*, 19(4), 875-894. <https://doi.org/10.1108/TG-06-2025-0187>
7. Andersen, L. B., Boesen, A., & Pedersen, L. H. (2016). Performance in public organizations: Clarifying the conceptual space. *Public Administration Review*, 76(6), 852-862. <https://doi.org/10.1111/puar.12578>
8. Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, 103(3), 411-423. <https://doi.org/10.1037/0033-2909.103.3.411>
9. Armstrong, J. S., & Overton, T. S. (1977). Estimating nonresponse bias in mail surveys. *Journal of Marketing Research*, 14(3), 396-402. <https://doi.org/10.1177/002224377701400320>
10. Bannister, F., & Connolly, R. (2014). ICT, public values and transformative government: A framework and programme for research. *Government Information Quarterly*, 31(1), 119-128. <https://doi.org/10.1016/j.giq.2013.06.002>
11. Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*, 25(24), 3186-3191. <https://doi.org/10.1097/00007632-200012150-00014>
12. Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing artificial intelligence. *MIS Quarterly*, 45(3), 1433-1450. <https://doi.org/10.25300/MISQ/2021/16274>
13. Boyne, G. A. (2003). Sources of public service improvement: A critical review and research agenda. *Journal of Public Administration Research and Theory*, 13(3), 367-394. <https://doi.org/10.1093/jopart/mug027>

15. Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology*, 1(3), 185-216. <https://doi.org/10.1177/135910457000100301>
16. Bullock, J. B. (2019). Artificial intelligence, discretion, and bureaucracy. *The American Review of Public Administration*, 49(7), 751-761. <https://doi.org/10.1177/0275074019856123>
17. Burton, J. W., Stein, M.-K., & Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making*, 33(2), 220-239. <https://doi.org/10.1002/bdm.2155>
18. Burton-Jones, A., & Straub, D. W. (2006). Reconceptualizing system usage: An approach and empirical test. *Information Systems Research*, 17(3), 228-246. <https://doi.org/10.1287/isre.1060.0096>
19. Busuioc, M. (2021). Accountable artificial intelligence: Holding algorithms to account. *Public Administration Review*, 81(5), 825-836. <https://doi.org/10.1111/puar.13293>
20. Cao, G., Duan, Y., & Li, G. (2015). Linking business analytics to decision making effectiveness: A path model analysis. *IEEE Transactions on Engineering Management*, 62(3), 384-395. <https://doi.org/10.1109/TEM.2015.2441875>
21. Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14(3), 464-504. <https://doi.org/10.1080/10705510701301834>
22. Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9(2), 233-255. https://doi.org/10.1207/S15328007SEM0902_5
23. Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum.
24. Cordella, A., & Tempini, N. (2015). E-government and organizational change: Reappraising the role of ICT and bureaucracy in public service delivery. *Government Information Quarterly*, 32(3), 279-286. <https://doi.org/10.1016/j.giq.2015.03.005>
25. DeLone, W. H., & McLean, E. R. (1992). Information systems success: The quest for the dependent variable. *Information Systems Research*, 3(1), 60-95. <https://doi.org/10.1287/isre.3.1.60>
26. DeLone, W. H., & McLean, E. R. (2003). The DeLone and McLean model of information systems success: A ten-year update. *Journal of Management Information Systems*, 19(4), 9-30. <https://doi.org/10.1080/07421222.2003.11045748>
27. Desouza, K. C., Dawson, G. S., & Chenok, D. (2020). Designing, developing, and deploying artificial intelligence systems: Lessons from and for the public sector. *Business Horizons*, 63(2), 205-213. <https://doi.org/10.1016/j.bushor.2019.11.004>
28. Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1), 114-126. <https://doi.org/10.1037/xge0000033>
29. Dishaw, M. T., & Strong, D. M. (1999). Extending the technology acceptance model with task-technology fit constructs. *Information & Management*, 36(1), 9-21. [https://doi.org/10.1016/S0378-7206\(98\)00101-3](https://doi.org/10.1016/S0378-7206(98)00101-3)
30. Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data - evolution, challenges and research agenda. *International Journal of Information Management*, 48, 63-71. <https://doi.org/10.1016/j.ijinfomgt.2019.01.021>
31. Dunleavy, P., Margetts, H., Bastow, S., & Tinkler, J. (2006). New public management is dead-Long live digital-era governance. *Journal of Public Administration Research and Theory*, 16(3), 467-494. <https://doi.org/10.1093/jopart/mui057>
32. Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., et al. (2021). Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, 57, 101994. <https://doi.org/10.1016/j.ijinfomgt.2019.08.002>
33. Floropoulos, J., Spathis, C., Halvatzis, D., & Tsiouridou, M. (2010). Measuring the success of the Greek taxation information system. *International Journal of Information Management*, 30(1), 47-56. <https://doi.org/10.1016/j.ijinfomgt.2009.03.013>
34. Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39-50. <https://doi.org/10.1177/002224378101800104>
35. Fügner, A., Grahl, J., Gupta, A., & Ketter, W. (2021). Will humans-in-the-loop become borgs? Merits and pitfalls of working with AI. *MIS Quarterly*, 45(3), 1527-1556. <https://doi.org/10.25300/MISQ/2021/16553>
36. Gable, G. G., Sedera, D., & Chan, T. (2008). Re-conceptualizing information system success: The IS-Impact Measurement Model. *Journal of the Association for Information Systems*, 9(7), 377-408. <https://doi.org/10.17705/1jais.00164>
37. Gaskin, J. E., Lowry, P. B., Rosengren, W., & Fife, P. T. (2025). Essential validation criteria for rigorous covariance-based structural equation modelling. *Information Systems Journal*, 35(6), 1630-1661. <https://doi.org/10.1111/isj.12598>
38. Ghasemaghahi, M., Ebrahimi, S., & Hassanein, K. (2018). Data analytics competency for improving firm decision making performance. *Journal of Strategic Information Systems*, 27(1), 101-113. <https://doi.org/10.1016/j.jsis.2017.10.001>

39. Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, 14(2), 627-660. <https://doi.org/10.5465/annals.2018.0057>
40. Goodhue, D. L. (1998). Development and measurement validity of a task-technology fit instrument for user evaluations of information system. *Decision Sciences*, 29(1), 105-138. <https://doi.org/10.1111/j.1540-5915.1998.tb01346.x>
41. Goodhue, D. L., & Thompson, R. L. (1995). Task-technology fit and individual performance. *MIS Quarterly*, 19(2), 213-236. <https://doi.org/10.2307/249689>
42. Gorla, N., Somers, T. M., & Wong, B. (2010). Organizational impact of system quality, information quality, and service quality. *Journal of Strategic Information Systems*, 19(3), 207-228. <https://doi.org/10.1016/j.jsis.2010.05.001>
43. Grimmelhuijsen, S. (2023). Explaining why the computer says no: Algorithmic transparency affects the perceived trustworthiness of automated decision-making. *Public Administration Review*, 83(2), 241-262. <https://doi.org/10.1111/puar.13483>
44. Haesevoets, T., De Cremer, D., Dierckx, K., & Van Hiel, A. (2021). Human-machine collaboration in managerial decision making. *Computers in Human Behavior*, 119, 106730. <https://doi.org/10.1016/j.chb.2021.106730>
45. Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, 43(1), 115-135. <https://doi.org/10.1007/s11747-014-0403-8>
46. Hu, L.-t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1-55. <https://doi.org/10.1080/10705519909540118>
47. Janssen, M., Hartog, M., Matheus, R., Yi Ding, A., & Kuk, G. (2022). Will algorithms blind people? The effect of explainable AI and decision-makers' experience on AI-supported decision-making in government. *Social Science Computer Review*, 40(2), 478-493. <https://doi.org/10.1177/0894439320980118>
48. Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons*, 61(4), 577-586. <https://doi.org/10.1016/j.bushor.2018.03.007>
49. Jasperson, J. S., Carter, P. E., & Zmud, R. W. (2005). A comprehensive conceptualization of post-adoptive behaviors associated with information technology enabled work systems. *MIS Quarterly*, 29(3), 525-557. <https://doi.org/10.2307/25148694>
50. Jeyaraj, A. (2020). DeLone & McLean models of information system success: Critical meta-review and research directions. *International Journal of Information Management*, 54, 102139. <https://doi.org/10.1016/j.ijinfomgt.2020.102139>
51. Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2022). semTools: Useful tools for structural equation modeling (R package version 0.5-6) [Computer software]. <https://CRAN.R-project.org/package=semTools>
52. Keding, C., & Meissner, P. (2021). Managerial overreliance on AI-augmented decision-making processes: How the use of AI-based advisory systems shapes choice behavior in R&D investment decisions. *Technological Forecasting and Social Change*, 171, 120970. <https://doi.org/10.1016/j.techfore.2021.120970>
53. Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.
54. Kock, N. (2015). Common method bias in PLS-SEM: A full collinearity assessment approach. *International Journal of e-Collaboration*, 11(4), 1-10. <https://doi.org/10.4018/ijec.2015100101>
55. Kuziemski, M., & Misuraca, G. (2020). AI governance in the public sector: Three tales from the frontiers of automated decision-making in democratic settings. *Telecommunications Policy*, 44(6), 101976. <https://doi.org/10.1016/j.telpol.2020.101976>
56. Lebovitz, S., Lifshitz-Assaf, H., & Levina, N. (2022). To engage or not to engage with AI for critical judgments: How professionals deal with opacity when using AI for medical diagnosis. *Organization Science*, 33(1), 126-148. <https://doi.org/10.1287/orsc.2021.1549>
57. Li, C.-H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, 48(3), 936-949. <https://doi.org/10.3758/s13428-015-0619-7>
58. Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, 151, 90-103. <https://doi.org/10.1016/j.obhdp.2018.12.005>
59. MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, 1(2), 130-149. <https://doi.org/10.1037/1082-989X.1.2.130>
60. MacKenzie, S. B., Podsakoff, P. M., & Podsakoff, N. P. (2011). Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques. *MIS Quarterly*, 35(2), 293-334. <https://doi.org/10.2307/23044045>

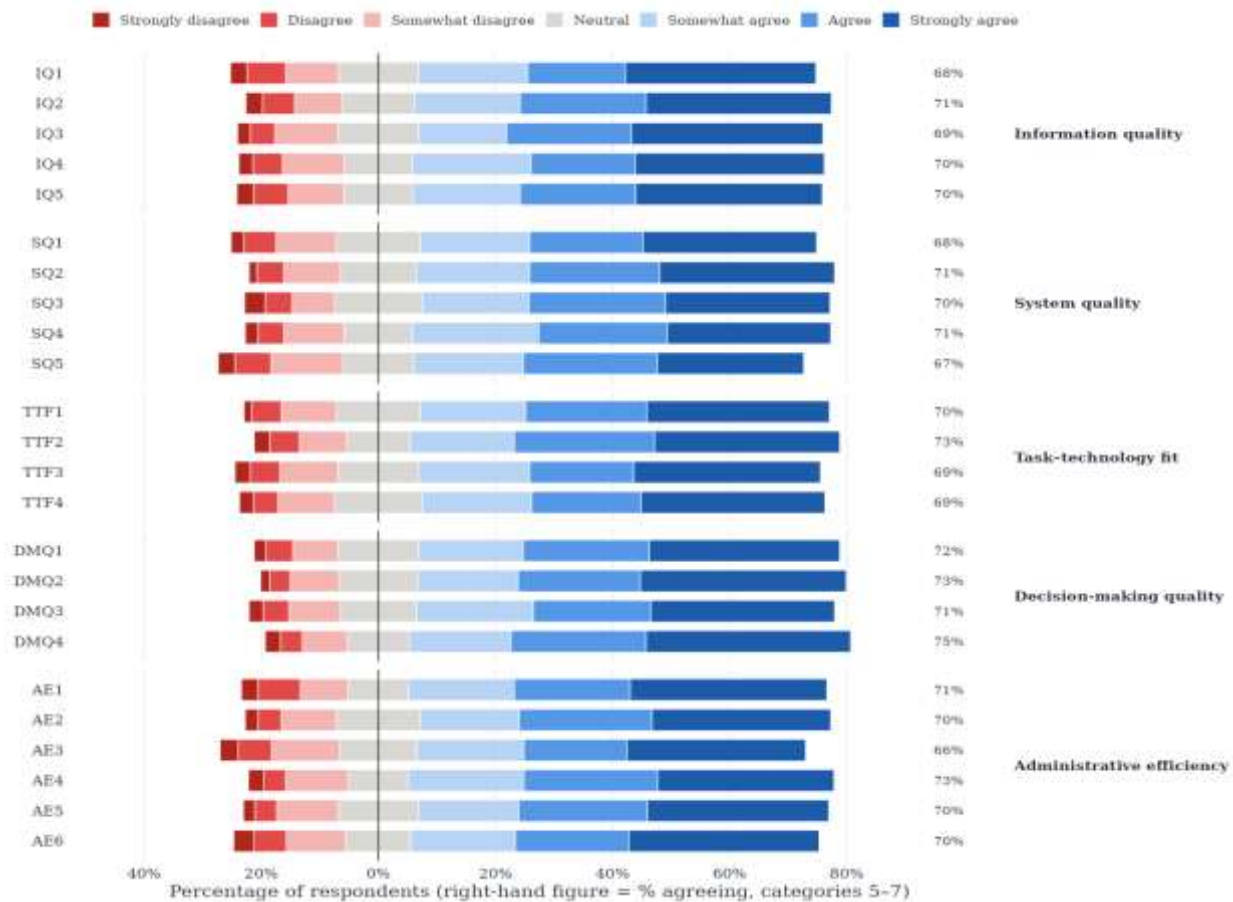
61. Madan, R., & Ashok, M. (2023). AI adoption and diffusion in public administration: A systematic literature review and future research agenda. *Government Information Quarterly*, 40(1), 101774. <https://doi.org/10.1016/j.giq.2022.101774>
62. Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57(3), 519-530. <https://doi.org/10.1093/biomet/57.3.519>
63. Medaglia, R., Gil-Garcia, J. R., & Pardo, T. A. (2023). Artificial intelligence in government: Taking stock and moving forward. *Social Science Computer Review*, 41(1), 123-140. <https://doi.org/10.1177/08944393211034087>
64. Melville, N., Kraemer, K., & Gurbaxani, V. (2004). Review: Information technology and organizational performance: An integrative model of IT business value. *MIS Quarterly*, 28(2), 283-322. <https://doi.org/10.2307/25148636>
65. Mikalef, P., Lemmer, K., Schaefer, C., Ylinen, M., Fjørtoft, S. O., Torvatn, H. Y., Gupta, M., & Niehaves, B. (2022). Enabling AI capabilities in government agencies: A study of determinants for European municipalities. *Government Information Quarterly*, 39(4), 101596. <https://doi.org/10.1016/j.giq.2021.101596>
66. Nelson, R. R., Todd, P. A., & Wixom, B. H. (2005). Antecedents of information and system quality: An empirical examination within the context of data warehousing. *Journal of Management Information Systems*, 21(4), 199-235. <https://doi.org/10.1080/07421222.2005.11045823>
67. Neumann, O., Guirguis, K., & Steiner, R. (2024). Exploring artificial intelligence adoption in public organizations: A comparative case study. *Public Management Review*, 26(1), 114-141. <https://doi.org/10.1080/14719037.2022.2048685>
68. Ouiddad, A., Okar, C., Chroqui, R., & Beqqali Hassani, I. (2021). Assessing the impact of enterprise resource planning on decision-making quality. *Kybernetes*, 50(5), 1144-1162. <https://doi.org/10.1108/K-04-2019-0273>
69. Pang, M.-S., Lee, G., & DeLone, W. H. (2014). IT resources, organizational capabilities, and value creation in public-sector organizations: A public-value management perspective. *Journal of Information Technology*, 29(3), 187-205. <https://doi.org/10.1057/jit.2014.2>
70. Petter, S., DeLone, W., & McLean, E. (2008). Measuring information systems success: Models, dimensions, measures, and interrelationships. *European Journal of Information Systems*, 17(3), 236-263. <https://doi.org/10.1057/ejis.2008.15>
71. Petter, S., DeLone, W., & McLean, E. R. (2013). Information systems success: The quest for the independent variables. *Journal of Management Information Systems*, 29(4), 7-62. <https://doi.org/10.2753/MIS0742-1222290401>
72. Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879-903. <https://doi.org/10.1037/0021-9010.88.5.879>
73. Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology*, 63, 539-569. <https://doi.org/10.1146/annurev-psych-120710-100452>
74. Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health*, 30(4), 459-467. <https://doi.org/10.1002/nur.20199>
75. Popović, A., Hackney, R., Coelho, P. S., & Jaklič, J. (2012). Towards business intelligence systems success: Effects of maturity and culture on analytical decision making. *Decision Support Systems*, 54(1), 729-739. <https://doi.org/10.1016/j.dss.2012.08.017>
76. Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, 40(3), 879-891. <https://doi.org/10.3758/BRM.40.3.879>
77. R Core Team. (2024). R: A language and environment for statistical computing (Version 4.3.3) [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org/>
78. Raghunathan, S. (1999). Impact of information quality and decision-maker quality on decision quality: A theoretical model and simulation analysis. *Decision Support Systems*, 26(4), 275-286. [https://doi.org/10.1016/S0167-9236\(99\)00060-3](https://doi.org/10.1016/S0167-9236(99)00060-3)
79. Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation-augmentation paradox. *Academy of Management Review*, 46(1), 192-210. <https://doi.org/10.5465/amr.2018.0072>
80. Richardson, H. A., Simmering, M. J., & Sturman, M. C. (2009). A tale of three perspectives: Examining post hoc statistical techniques for detection and correction of common method variance. *Organizational Research Methods*, 12(4), 762-800. <https://doi.org/10.1177/1094428109332834>
81. Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1-36. <https://doi.org/10.18637/jss.v048.i02>
82. Sabherwal, R., Jeyaraj, A., & Chowa, C. (2006). Information system success: Individual and organizational determinants. *Management Science*, 52(12), 1849-1864. <https://doi.org/10.1287/mnsc.1060.0583>
83. Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, 66(4), 507-514. <https://doi.org/10.1007/BF02296192>

84. Saudi Data and AI Authority. (2020). National strategy for data & AI. SDAIA. https://wp.oecd.ai/app/uploads/2021/12/Saudi_Arabia_National_Strategy_for_Data_and_AI_2020.pdf
85. Seddon, P. B. (1997). A respecification and extension of the DeLone and McLean model of IS success. *Information Systems Research*, 8(3), 240-253. <https://doi.org/10.1287/isre.8.3.240>
86. Selten, F., Robeer, M., & Grimmelikhuijsen, S. (2023). 'Just like I thought': Street-level bureaucrats trust AI recommendations if they confirm their professional judgment. *Public Administration Review*, 83(2), 263-278. <https://doi.org/10.1111/puar.13602>
87. Shrestha, Y. R., Ben-Menahem, S. M., & von Krogh, G. (2019). Organizational decision-making structures in the age of artificial intelligence. *California Management Review*, 61(4), 66-83. <https://doi.org/10.1177/0008125619862257>
88. Sun, T. Q., & Medaglia, R. (2019). Mapping the challenges of artificial intelligence in the public sector: Evidence from public healthcare. *Government Information Quarterly*, 36(2), 368-383. <https://doi.org/10.1016/j.giq.2018.09.008>
89. Todd, P., & Benbasat, I. (1999). Evaluating the impact of DSS, cognitive effort, and incentives on strategy selection. *Information Systems Research*, 10(4), 356-374. <https://doi.org/10.1287/isre.10.4.356>
90. Torkzadeh, G., & Doll, W. J. (1999). The development of a tool for measuring the perceived impact of information technology on work. *Omega*, 27(3), 327-339. [https://doi.org/10.1016/S0305-0483\(98\)00049-8](https://doi.org/10.1016/S0305-0483(98)00049-8)
91. Trunk, A., Birkel, H., & Hartmann, E. (2020). On the current state of combining human and artificial intelligence for strategic organizational decision making. *Business Research*, 13(3), 875-919. <https://doi.org/10.1007/s40685-020-00133-x>
92. Twizeyimana, J. D., & Andersson, A. (2019). The public value of e-government - A literature review. *Government Information Quarterly*, 36(2), 167-178. <https://doi.org/10.1016/j.giq.2019.01.001>
93. van Noordt, C., & Misuraca, G. (2022). Artificial intelligence for the public sector: Results of landscaping the use of AI in government across the European Union. *Government Information Quarterly*, 39(3), 101714. <https://doi.org/10.1016/j.giq.2022.101714>
94. Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3(1), 4-70. <https://doi.org/10.1177/109442810031002>
95. Westland, J. C. (2010). Lower bounds on sample size in structural equation modeling. *Electronic Commerce Research and Applications*, 9(6), 476-487. <https://doi.org/10.1016/j.elerap.2010.07.003>
96. Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2019). Artificial intelligence and the public sector-Applications and challenges. *International Journal of Public Administration*, 42(7), 596-615. <https://doi.org/10.1080/01900692.2018.1498103>
97. Wirtz, B. W., Weyerer, J. C., & Kehl, I. (2022). Governance of artificial intelligence: A risk and guideline-based integrative framework. *Government Information Quarterly*, 39(4), 101685. <https://doi.org/10.1016/j.giq.2022.101685>
98. Wixom, B. H., & Todd, P. A. (2005). A theoretical integration of user satisfaction and technology acceptance. *Information Systems Research*, 16(1), 85-102. <https://doi.org/10.1287/isre.1050.0042>
99. Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement*, 73(6), 913-934. <https://doi.org/10.1177/0013164413495237>
100. Zhao, X., Lynch, J. G., Jr., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of Consumer Research*, 37(2), 197-206. <https://doi.org/10.1086/651257>
101. Ziguers, I., & Buckland, B. K. (1998). A theory of task/technology fit and group support systems effectiveness. *MIS Quarterly*, 22(3), 313-334. <https://doi.org/10.2307/249668>
102. Zuiderwijk, A., Chen, Y.-C., & Salem, F. (2021). Implications of the use of artificial intelligence in public governance: A systematic literature review and a research agenda. *Government Information Quarterly*, 38(3), 101577. <https://doi.org/10.1016/j.giq.2021.101577>

Appendix A. Supplementary Figures

Figure A1

Item-Level Response Distributions for the 24 Measurement Items



Note. N = 455. Diverging stacked bars centred on the neutral midpoint (category 4, split evenly either side of zero). Disagreement categories extend to the left and agreement categories to the right; the right-hand figure gives the percentage of respondents choosing categories 5-7.

Figure A2. Discriminant Validity: Latent Correlations, $\sqrt{\text{AVE}}$ and HTMT Ratios



Note. Lower triangle: latent (CFA) correlations; diagonal (bold): square root of AVE; upper triangle (italic): HTMT ratios. Darker cells indicate higher values. Values correspond to Table 10.