Beyond Trust: Designing a Perceived Intelligence Scale for AI-Powered Decision Support Systems in the Public Sector

Muhammad Shafeeq Mohd Sapian[1*], Haslinda Musa[1], Nurulizwa Rashid[1], Fadhlur Rahim Azmi[2], and Muhammad Shafiq Ibrahim[3]

[1]Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, 76100 Melaka, Malaysia

[2]Faculty of Business and Management, Universiti Teknologi MARA, Melaka, Malaysia

[3]Department of Materials, Manufacturing and Industrial Engineering, School of Mechanical Engineering, Faculty of Engineering, Universiti Teknologi Malaysia, Johor Bahru, Malaysia

[*]Corresponding Author

DOI: https://dx.doi.org/10.47772/IJRISS.2026.101000821

Received: 25 October 2026; Accepted: 30 October 2026; Published: 19 November 2026

ABSTRACT

Artificial Intelligence-Powered Decision Support Systems (AI-DSS) are spreading unevenly across public-sector agencies, and instruments built on the Unified Theory of Acceptance and Use of Technology (UTAUT) were not designed to capture how users judge an AI system's intelligence. This paper reports on the design of a Perceived Intelligence (PI) scale for AI-DSS adoption research. Five dimensions, namely predictive accuracy, adaptive learning, contextual awareness, explanatory transparency, and overall intelligence, were specified from the literature on perceived intelligence, trust, and explainable AI, and ten items were drafted on a five-point Likert scale. Semi-structured interview probes have been prepared for officers in Malaysia's health, education, and finance ministries to confirm or refine these dimensions under decision rules fixed in advance. The paper also sets out the content-validation protocol to follow, with item retention based on an item-level content validity index (I-CVI) of at least .78 and a scale-level index (S-CVI/Ave) of at least .90. No interview, expert-review, or pilot data have been collected yet. What the paper contributes is the item pool itself and a validation protocol fixed before fieldwork begins.

Keywords: Perceived intelligence; trust; AI-powered decision support systems; scale development; UTAUT

INTRODUCTION

Governments across the world are turning to Artificial Intelligence-Powered Decision Support Systems (AI-DSS) to speed up policy formation, service delivery, and resource allocation (United Nations Department of Economic and Social Affairs, 2022; World Bank, 2021). Malaysia's own digital-government agenda, spanning the Public Sector Digitalisation Strategic Plan 2021–2025, the National Artificial Intelligence Roadmap 2021–2025, and the Twelfth Malaysia Plan, pushes ministries such as Health, Education, and Finance toward the same goal (Economic Planning Unit, 2021; MAMPU, 2021; Ministry of Science, Technology and Innovation, 2021). Adoption nonetheless meets persistent obstacles, among them a shortage of qualified expertise, concerns over data privacy, and resistance to organisational change (Mahusin et al., 2024). Systematic reviews of AI adoption in public-sector governance identify a similar mix of determinants and challenges (Sapian et al., 2026d), and a cross-sectoral policy review of Malaysia's health, education, and finance ministries finds that their readiness for AI-DSS differs (Sapian et al., 2026a). A further reason lies in the theories used to explain adoption, which were developed for conventional information systems rather than for systems that learn, predict, and increasingly share decision work with their users (Rai et al., 2019; Shin, 2021).

These systems are no longer hypothetical in Malaysia. The Ministry of Health has introduced AI-assisted chest X-ray screening at selected public health clinics and DR.MATA for the detection of diabetic retinopathy (Al As, 2025; CodeBlue, 2025). The Ministry of Education has strengthened its Student Tracking System (SiPKPM) with AI to identify pupils at risk of dropping out (Bernama, 2025). Within the Ministry of Finance

portfolio, customs officers work with AI-enabled baggage scanners at entry points, and the Inland Revenue Board selects audit cases through computerised risk analysis (Bernama, 2024; Inland Revenue Board of Malaysia, 2025). In each of these settings an officer receives the system's output and decides whether to act on it. That judgement, and the officer's sense of how capable the system is, is what a measure of perceived intelligence has to capture.

The Unified Theory of Acceptance and Use of Technology (UTAUT) is the dominant lens for studying public-sector technology adoption (Venkatesh et al., 2003), and it has been extended many times since (Dwivedi et al., 2019; Tamilmani et al., 2021). What these extensions have not yet captured consistently is how a user judges an AI system's intelligence: whether its predictions are trustworthy, whether it adapts, whether it understands the context it operates in, and whether it can explain itself. Recent systematic reviews of UTAUT extensions for AI and AI-DSS adoption (Sapian et al., 2026b, 2026c) point to the same gap. Moussawi and Koufaris (2019) developed and validated a perceived-intelligence scale for personal intelligent agents, and Moussawi et al. (2021) showed that perceived intelligence is a significant antecedent of their adoption. That work concerns consumer assistants, however, not decision support in government. Shin (2021) makes a related point about explainable AI: users do not extend trust to an algorithm just because it works. They extend it because they can follow, at some level, why it produced the output it did.

The present paper addresses this gap at the level of measurement rather than through an adoption study. We report the design of Perceived Intelligence (PI), a five-dimension, ten-item scale intended to sit inside a UTAUT-based AI-DSS adoption model as a mediating construct and to be refined through fieldwork with ministry personnel in Malaysia's health, education, and finance sectors. The sections that follow trace how the item pool was derived from the literature and describe the interview probes that will confirm or refine it. They also state plainly what has and has not yet been done to validate the scale, and set out the protocol for the content-validation step that comes next. Since data collection has not begun, no reliability or validity statistics are reported. The contribution lies in the item pool, the theoretical rationale for each dimension, and a validation plan specified in enough detail to be replicated.

The paper therefore sets out to define Perceived Intelligence and to draw its conceptual boundaries with performance expectancy, effort expectancy, and trust. It then presents a theory-grounded pool of ten items across five a priori dimensions. Finally, it specifies, before any data are collected, the qualitative refinement, content-validation, translation, and pilot procedures through which the scale will be validated.

THEORETICAL GROUNDING OF PERCEIVED INTELLIGENCE

Perceived Intelligence, as used here, is the user's perception of an AI-DSS's capacity to learn, reason, predict accurately, read context, and explain itself (Moussawi & Koufaris, 2019; Moussawi et al., 2021). Several bodies of work point toward this idea, although none names it in quite the same way. The most direct evidence comes from research on intelligent agents, where perceived intelligence is a significant antecedent of adoption (Moussawi et al., 2021) and is central to continued use (Moussawi et al., 2023). Trust research approaches the same territory from another side. Gefen et al. (2003) showed that trust operates alongside perceived usefulness and ease of use in shaping intention, and McKnight et al. (2002) decomposed trusting beliefs into competence, benevolence, and integrity, a decomposition that maps reasonably well onto the dimensions proposed here. Explainable AI (XAI) research supplies the transparency side of the construct: Shin (2021) found that explainability and causability shape how users perceive, trust, and accept algorithmic output. Technology-acceptance scholars, for their part, have been calling for AI-specific extensions to TAM and UTAUT for several years (Dwivedi et al., 2019), on the grounds that ease of use and usefulness say little about what is distinctive in a system that adapts.

Table 1 summarises how each body of work informs the construct.

Table 1. Theoretical grounding of Perceived Intelligence

Literature	Contribution to Perceived Intelligence	Source
Trust theory	Trust works alongside perceived usefulness and ease of use in	Gefen et al. (2003);

	shaping intention; explanation shapes trust in AI	Shin (2021)
Intelligent-agent research	Perceived intelligence is a significant antecedent of the adoption and continued use of intelligent agents	Moussawi et al. (2021, 2023)
Explainable AI (XAI)	Explainability and causability shape how algorithmic output is perceived and accepted	Shin (2021)
Technology-acceptance research	AI-specific constructs are needed to extend models such as TAM and UTAUT	Dwivedi et al. (2019)
Trust beliefs	Trusting beliefs comprise competence, benevolence, and integrity	McKnight et al. (2002)

None of these literatures, on its own, supplies a ready-made measurement scale for AI-DSS specifically. The perceived-intelligence scale of Moussawi and Koufaris (2019) is the closest precedent and informed the wording of three items; the remaining items were written for the ministry AI-DSS setting and will be checked against the qualitative phase described next.

Measures developed in consumer settings do not transfer directly to public-sector decision support. A consumer who finds a voice assistant unhelpful can stop using it at will, whereas an officer is often required to use a ministry system and remains accountable for the decision it informs. Officers also judge a system against institutional constraints that consumers never face, such as statutory procedures, ministry-specific data, and the expectations of superiors. These differences are the reason Contextual Awareness is treated as a separate dimension, and the reason every PI item refers to the specific AI-DSS that the respondent uses or whose outputs reach their work, rather than to artificial intelligence in general.

Malaysian adoption research has, for its part, concentrated on organisation-level factors such as the characteristics of an innovation, organisational readiness, and external pressures (Azmi et al., 2020; Musa et al., 2016). These studies explain whether an organisation adopts a standard or system, but not how an individual officer judges the capability of the system they work with, which is the judgement the PI scale is designed to capture.

Because the construct sits inside a UTAUT-based model, its boundaries with the neighbouring constructs need to be explicit. Performance expectancy concerns what using the system does for the officer's work; effort expectancy concerns how easy it is to use; Perceived Intelligence concerns the officer's judgement of the system's own capability. Trust is a willingness to rely on the system when outcomes are uncertain. It is treated as a theoretical consequence of Perceived Intelligence and is not measured separately, so that the PI items capture judged capability, not willingness to rely on the system. Perceived Intelligence also excludes judgements about a system's benevolence or the legitimacy of the institution that deploys it, which belong to trust rather than to perceived capability (McKnight et al., 2002); none of the five dimensions asks whether the system acts in the officer's or the public's interest. Table 2 sets out these distinctions as they are defined in the parent study.

Table 2. Conceptual boundaries between Perceived Intelligence and neighbouring constructs

Construct	**Definition in the parent study**	**Dimensions**	**Role in the parent model**
Performance Expectancy (PE)	The degree to which an officer believes that using the AI-DSS improves job performance and decision outcomes	Unidimensional	Independent variable
Effort Expectancy (EE)	The degree of ease associated with using the AI-DSS	Unidimensional	Independent variable
Perceived Intelligence (PI)	An officer's judgement that the AI-DSS demonstrates capable, adaptive, context-aware, and explainable behaviour; distinct from trust	Predictive accuracy; adaptive learning; contextual awareness; explanatory transparency;	Mediator between PE/EE and behavioural intention

		overall intelligence	
Trust	A willingness to rely on the AI-DSS when outcomes are uncertain	Not measured	Theoretical consequence of PI

Figure 1 shows where the scale sits in the parent study. The parent study extends UTAUT with Perceived Intelligence as a mediator and Governance Efficiency as an outcome, and tests twelve hypotheses. PI mediates the effects of performance expectancy and effort expectancy on behavioural intention (H7 and H8, through H5); gender, organisational grade or role (measured by service group), and education moderate selected paths (H10–H12); and adoption or use is linked to perceived governance efficiency (H9). The shaded region marks the part of the model this paper develops: the PI construct and its five reflectively measured dimensions. Every hypothesis that involves PI (H5, H7, H8, and H10) depends on the scale being valid, which is why the scale is developed and validated before the structural model is tested.

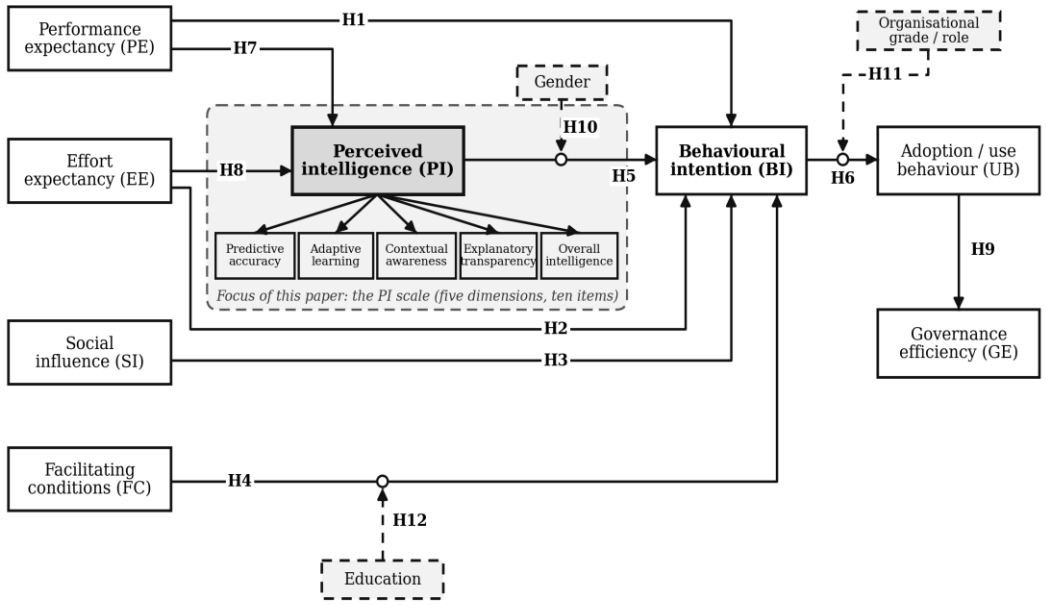


Figure 1. Position of the Perceived Intelligence scale within the parent study's extended UTAUT framework

Note: Solid arrows show hypothesised direct and mediated paths (H1–H9); dashed arrows ending at a circle show moderating effects (H10–H12). Arrows from PI to its five dimensions indicate the reflective measurement developed in this paper. Adapted from the conceptual framework of the parent study.

Table 3 lists the twelve hypotheses of the parent study and the main sources on which each rests. They are reported here for context; this paper tests none of them.

Table 3. Hypotheses of the parent study and supporting sources

H	Path	Hypothesis	Supporting sources
H1	PE → BI	Performance expectancy positively influences behavioural intention to use AI-DSS	Venkatesh et al. (2003)
H2	EE → BI	Effort expectancy positively influences behavioural intention	Venkatesh et al. (2003)
H3	SI → BI	Social influence positively influences behavioural intention	Venkatesh et al. (2003); Mahusin et al. (2024)
H4	FC → BI	Facilitating conditions positively influence behavioural intention	Venkatesh et al. (2003); Mahusin et al. (2024)
H5	PI → BI	Perceived intelligence positively influences behavioural intention	Moussawi et al. (2021, 2023); Shin (2021)
H6	BI → UB	Behavioural intention positively influences adoption/use	Ajzen (1991); Venkatesh et al.

		behaviour	(2003)
H7	PE → PI → BI	Perceived intelligence mediates the relationship between performance expectancy and behavioural intention	Venkatesh et al. (2003); Moussawi et al. (2021)
H8	EE → PI → BI	Perceived intelligence mediates the relationship between effort expectancy and behavioural intention	Venkatesh et al. (2003); Shin (2021)
H9	UB → GE	Adoption/use of AI-DSS is positively associated with perceived governance efficiency	Twizeyimana & Andersson (2019); Medrano-Sánchez & Jara (2026)
H10	Gender × (PI → BI)	Gender moderates the relationship between perceived intelligence and behavioural intention	Venkatesh & Morris (2000)
H11	Grade/role × (BI → UB)	Organisational grade/role (service group) moderates the relationship between behavioural intention and adoption/use	Venkatesh et al. (2003)
H12	Education × (FC → BI)	Education moderates the relationship between facilitating conditions and behavioural intention	Nawafleh & Fares (2024)

METHODOLOGY

We followed the scale-development sequence of DeVellis (2017): define the construct, generate an item pool, decide the response format, and submit the pool to expert review before any pilot or main-sample data collection. This paper reports the first three stages and specifies in detail how the fourth will proceed.

Figure 2 summarises the five-stage development and validation process planned for the parent study, from initial item generation through expert content validation, translation, and pilot testing to the final bilingual instrument.

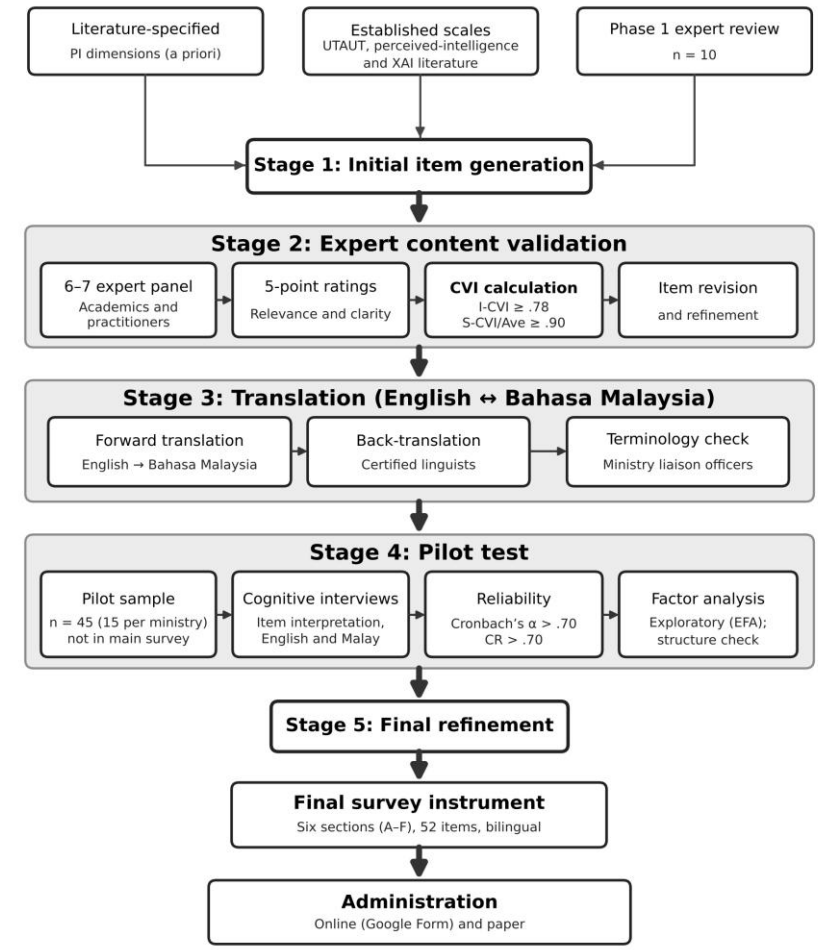


Figure 2. Development and validation process for the Perceived Intelligence scale

3.1 Qualitative refinement phase (planned)

The parent study's semi-structured interview guide contains nine core questions and is designed for interviews of 15–20 minutes with 20–25 officers selected purposively across the three ministries, their administrative tiers and the three service groups of the Malaysian public service (Top Management, Management and Professional, and Implementation). Two of these questions address Perceived Intelligence directly (Table 4). Q5 asks what makes the system seem intelligent or capable, and Q6 asks whether an intelligent system is the same as one the officer trusts. The five sub-dimensions of Perceived Intelligence (Predictive Accuracy, Adaptive Learning, Contextual Awareness, Explanatory Transparency, and Overall Intelligence) were specified in advance from the literature reviewed in Section 2, and Q5 is designed to test them. The open question is asked, and the participant's own answer recorded, before any dimension is offered as a probe, so that dimensions the framework did not anticipate can surface.

Responses will be coded thematically under decision rules fixed before the interviews take place. An a priori dimension is retained if participants in at least two ministries describe it; two dimensions are merged if participants consistently treat them as one; and a new dimension is added if it recurs across at least two ministries and is not captured by existing items. Q6 serves a different purpose: it tests whether officers themselves separate a capable system from a trusted one, which is the boundary on which the construct rests. Grounding items in the target population's own understanding in this way is standard practice in scale development (Clark & Watson, 1995; DeVellis, 2017; Lynn, 1986).

Transcripts will be analysed in NVivo 14 using both deductive and inductive coding: a priori codes derived from the UTAUT framework and the five PI dimensions give initial structure, while emergent codes capture unexpected themes. Coding proceeds from open coding of relevant segments, through axial coding into categories, to selective coding into higher-order themes. Two researchers will independently code 20% of the transcripts, with Cohen's kappa of at least .80 as the agreement target, and discrepancies will be resolved jointly to refine the codebook. Themes will then be compared across ministries, organisational levels, and functional roles.

Table 4. Interview questions on Perceived Intelligence in the parent study’s interview guide

Question	Optional probes	Purpose
Q5. What makes the system seem “intelligent” or capable to you?	Accuracy and consistency? Learning or adapting over time? Understanding your ministry’s context? Explaining its recommendations?	Identify the dimensions of Perceived Intelligence
Q6. Is an intelligent system the same as a system you trust? When do you accept, check or override its recommendations?	Can a system be capable but not trusted? Who is accountable when you follow it?	Distinguish perceived intelligence from trust

3.2 Item pool and response format

We drafted ten items across the five dimensions (Table 5). Three were informed by Moussawi and Koufaris (2019), two were adapted from Shin (2021), and the remaining five were written for the ministry AI-DSS setting and checked against the theoretical grounding above. All items use a five-point Likert format (1 = strongly disagree, 5 = strongly agree), matching the response format used for the UTAUT constructs elsewhere in the parent study, which allows the PI items to sit alongside them in a single survey instrument.

3.3 Planned content-validation protocol

Content validity will be assessed in Phase 3 of the parent study by a content-validity panel of six or seven experts drawn from the expert pool identified for the study, spanning technology-adoption research, public administration, artificial intelligence/machine-learning practice, and domain expertise across health, education, and finance. Each expert will rate every item for relevance to its intended construct and for clarity of wording

on a five-point scale (1 = strongly disagree/poor, 5 = strongly agree/excellent), using Part 2 of the parent study's expert validation form. The I-CVI is calculated from the relevance ratings as the proportion of experts who rate an item 4 or 5. Items with an I-CVI below .78 will be revised or removed (Lynn, 1986), and each scale must reach an S-CVI/Ave of at least .90 (Polit & Beck, 2006). Six or seven experts are used because Lynn (1986) requires full agreement when five or fewer experts rate an item, so that the .78 criterion can apply as stated. No expert has yet completed a review.

3.4 Translation and pilot testing (planned)

The PI items will be administered within the parent study's bilingual survey instrument, which comprises six sections and 52 items. After content validation, the items will be translated from English into Bahasa Malaysia, back-translated by certified linguists, and checked by ministry liaison officers for terminology and cultural appropriateness. The instrument will then be piloted with 45 officers, 15 from each ministry, who will be excluded from the main survey. The pilot combines cognitive interviews on the English and Malay versions, to check how officers interpret each item, with reliability analysis (Cronbach's alpha and composite reliability above .70) and exploratory factor analysis to check the dimensional structure. Exploratory factor analysis of five-point perception items has been used for a similar purpose in Malaysian adoption research (Azmi et al., 2020). Items will be refined in light of these results before the final instrument is administered online through Google Forms and on paper.

3.5 Planned psychometric testing (future work)

In the main survey, Perceived Intelligence is specified as a reflective, multi-dimensional construct measured on a five-point Likert scale, and the items will be assessed alongside the rest of the instrument using Partial Least Squares Structural Equation Modelling. Indicator loadings are expected to exceed .708. Reliability will be judged against Cronbach's alpha and composite reliability thresholds above .70 (Hair et al., 2019, 2021; Nunnally & Bernstein, 1994); convergent validity against an average variance extracted above .50 (Fornell & Larcker, 1981); and discriminant validity against the Fornell-Larcker criterion and a heterotrait-monotrait ratio below .85 (Henseler et al., 2015). Discriminant validity will be examined at two levels: among the five PI dimensions, so that any pair that proves empirically redundant can be merged under the decision rules in Section 3.1, and between PI and its closest neighbouring constructs, performance expectancy and effort expectancy, to confirm that the scale captures judged capability as distinct from expected performance or ease of use. The incremental contribution of PI will be assessed by comparing the variance explained in behavioural intention with and without PI, using the change in R^2 and the f^2 effect size (Hair et al., 2021). Criterion-related validity will be examined through PI's hypothesised associations with behavioural intention and, through intention, with self-reported use (H5, H7, and H8). Because the main survey draws on officers at federal, state, and district levels in all three ministries, measurement invariance across ministries, tiers and service groups will be tested before scores are compared between groups (Henseler et al., 2016). Together these tests address the convergent and discriminant evidence that construct validity requires (Cronbach & Meehl, 1955). No such figures exist yet for the PI scale. The thresholds above are criteria for future data, not results, and we state them now so that the eventual validation study can be assessed against a protocol fixed before any data were collected.

THE PERCEIVED INTELLIGENCE ITEM POOL

Table 5 sets out the ten items across five dimensions. Predictive Accuracy (three items) and Adaptive Learning (two items) draw most directly on the perceived-intelligence scale of Moussawi and Koufaris (2019), adapted from a personal-assistant setting to a ministry AI-DSS setting. Contextual Awareness (two items) has no direct precedent in that literature. It was written to reflect the ministry-specific constraints that a system built for e-commerce or customer service would be likely to miss, and the Q5 probe on understanding the ministry's context is designed to test whether officers see it the same way. Explanatory Transparency (two items) draws on the XAI literature (Shin, 2021). Overall Intelligence (one item) is not a distinct facet. We included it as a global check, so that a single summary judgement can be read alongside the four facet scores and compared with them.

Table 5. Perceived Intelligence item pool by dimension

Code	Dimension	Item statement	Source
PI1	Predictive Accuracy	The AI-DSS provides accurate predictions and recommendations.	Informed by Moussawi & Koufaris (2019)
PI2	Predictive Accuracy	The AI-DSS makes reliable forecasts based on available data.	Developed for this study; grounded in McKnight et al. (2002)
PI3	Predictive Accuracy	The AI-DSS provides recommendations that remain consistent across similar situations.	Adapted from Shin (2021); trust framing removed
PI4	Adaptive Learning	The AI-DSS demonstrates an ability to learn from past data and interactions.	Informed by Moussawi & Koufaris (2019)
PI5	Adaptive Learning	The AI-DSS adapts its suggestions based on user interactions and feedback.	Developed for this study; grounded in Moussawi et al. (2021)
PI6	Contextual Awareness	The AI-DSS considers the specific context of my ministry.	Developed for this study; no direct precedent (see Section 2)
PI7	Contextual Awareness	The AI-DSS understands the constraints of public-sector decision-making.	Developed for this study; no direct precedent (see Section 2)
PI8	Explanatory Transparency	The AI-DSS explains its reasoning in understandable terms.	Adapted from Shin (2021)
PI9	Explanatory Transparency	I understand how the AI-DSS arrives at its recommendations.	Developed for this study; grounded in Shin (2021)
PI10	Overall Intelligence	Overall, I perceive the AI-DSS as an intelligent and capable system.	Informed by Moussawi & Koufaris (2019) (global check-item)

One feature of the item pool requires comment. PI3 was originally adapted from a trust item in Shin (2021), "I trust the recommendations provided by the AI-DSS". It has been reworded to remove the trust framing, so that the item now assesses the consistency of the system's recommendations and leaves willingness to rely on the system to the separate construct of trust.

The allocation of items to dimensions follows the probes attached to interview question Q5. Predictive Accuracy carries three items because officers' first test of a system is usually whether its outputs are right and whether they stay right across similar cases; PI3 was written to capture the second of these. Adaptive Learning carries two items because the claim that an AI-DSS improves as it learns is part of what distinguishes it from conventional decision support. Contextual Awareness carries two items, one on the ministry's specific context and one on the wider constraints of public-sector decision-making; Explanatory Transparency carries two, one on the system's explanation and one on the officer's own understanding of it. If the interviews show that officers treat any two dimensions as one, the decision rules in Section 3.1 allow them to be merged before the content-validation round.

DISCUSSION

A dedicated PI instrument is needed because a general measure of trust or usefulness conceals the distinctions that matter most when officers evaluate an AI system. A ministry officer might trust an AI-DSS's fraud-detection output because it is usually accurate, while distrusting its transparency because the model cannot explain why a particular transaction was flagged. A single-item trust measure collapses that distinction, whereas PI's five sub-dimensions are built to preserve it. Q6 of the Phase 2 interviews asks officers directly whether a capable system is necessarily a trusted one, so the interviews will show whether officers draw this distinction themselves or whether the instrument imposes it.

Beyond research, the dimensions have diagnostic value for ministries. Low scores on Explanatory Transparency would point to interface and explanation design, while low scores on Contextual Awareness would suggest that a system configured centrally fits poorly with the work of a particular tier or unit. Because every item refers to a named system, scores can be compared across systems within the same ministry and tracked as a system matures from pilot to routine use.

The scale is equally relevant to Malaysia's TVET 2.0 agenda, under which the government is restructuring technical and vocational education and training to strengthen high-technology fields (Bernama, 2026b), with a Strategic TVET 2.0 Task Force set up for 2026–2035 (Bernama, 2026a). As AI-DSS spread through public agencies, upskilling and reskilling programmes for public servants could use the Explanatory Transparency and Contextual Awareness dimensions to identify where officers need support in judging what a system can and cannot do. Applying the scale to officers in TVET agencies, and to TVET instructors who work with AI-based systems, would show whether its dimensions hold beyond the three ministries studied here.

At this stage the scale rests on limited evidence. Content validity has not been tested: the expert panel has been identified, but no expert has yet completed a review. The Phase 2 interviews have not yet been conducted, so the dimension structure remains a literature-based specification. No pilot data exist. The instrument is intended for administration in both English and Bahasa Malaysia, but the translation-back-translation procedure used elsewhere in the parent study has not yet been applied specifically to the PI items. And the item pool, however carefully grounded, will be refined using a single-country, three-ministry qualitative sample. Whether Contextual Awareness generalises beyond Malaysian public-sector governance, or reflects something more local to it, is an open empirical question this paper cannot settle. Two further forms of validation fall outside the scope of the parent study. The scale is not tested against established measures of trust or against automation bias, the tendency to over-rely on automated recommendations (Parasuraman & Manzey, 2010), and its predictive validity rests on self-reported intention and use rather than on observed behaviour. Future studies should test whether PI scores predict behavioural outcomes such as the frequency with which officers override AI recommendations, accuracy in simulated decision tasks, or recorded decision accuracy, preferably in longitudinal designs.

What the paper does provide is a fixed reference point for the work that follows: a defined item pool, decision rules for the interviews, and pass/fail thresholds set before any data could influence them. In our view, fixing the protocol before data collection is as important a contribution as the ten items themselves.

CONCLUSION

This paper has reported the design of Perceived Intelligence, a five-dimension, ten-item scale intended to measure how users of AI-Powered Decision Support Systems judge those systems' intelligence: their accuracy, their capacity to adapt, their grasp of institutional context, their willingness to explain themselves, and a general sense of capability. The scale was specified from the perceived-intelligence, trust, and explainable-AI literature and is designed to be refined through interviews with Malaysian ministry personnel in health, education, and finance. Before the instrument can be used with any confidence, the Phase 2 interviews must be completed, the items must pass expert content review against the pre-specified I-CVI and S-CVI thresholds, and the scale must be piloted ahead of the confirmatory PLS-SEM testing that a subsequent paper will report. Each of these steps is specified here in replicable detail; none has yet been carried out. Future work should report the interview findings and the content-validation round explicitly, including any items dropped or revised, before proceeding to the substantive adoption model this instrument was built to serve.

ACKNOWLEDGEMENT

The authors thank the Ministry of Higher Education (MOHE) Malaysia for sponsoring this work through the Fundamental Research Grant Scheme FRGS/1/2025/SS01/UTEM/02/1 and the Faculty of Technology Management and Technopreneurship, Universiti Teknikal Malaysia Melaka (UTeM), for their support.

REFERENCES

1. Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T
2. Al As, M. (2025, May 4). Malaysia launches AI-powered X-ray to detect lung diseases. *New Straits Times*. https://www.nst.com.my/news/nation/2025/05/1211281/malaysia-launches-ai-powered-x-ray-detect-lung-diseases

3. Azmi, F. R., Abdullah, A., Musa, H., & Wan Mahmood, W. H. (2020). Perception of food manufacturers towards adoption of halal food supply chain in Malaysia: Exploratory factor analysis. *Journal of Islamic Marketing, 11*(3), 571–589. https://doi.org/10.1108/JIMA-12-2018-0236
4. Bernama. (2024, September 27). JKDM boosts security at entry points with installation of 40 AI scanners. *Bernama*. https://bernama.com/en/news.php?id=2345345
5. Bernama. (2025, February 14). Pengukuhan Sistem Pengesanan Murid dengan AI mampu atasi isu keciciran pelajar [Strengthening the Student Tracking System with AI can address student dropout]. *Bernama*. https://bernama.com/bm/news.php?id=2392485
6. Bernama. (2026a, August 6). Malaysia bolsters TVET agenda with TVET 2.0 strategic task force – Ahmad Zahid. *Bernama*. https://bernama.com/en/news.php?id=2591232
7. Bernama. (2026b, May 16). TVET 2.0 to prioritise high-impact areas, says Zahid. *New Straits Times*. https://www.nst.com.my/news/nation/2026/05/1441002/tvet-20-prioritise-high-impact-areas-says-zahid
8. Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*(3), 309–319. https://doi.org/10.1037/1040-3590.7.3.309
9. CodeBlue. (2025, March 28). MOH exploring AI for cancer, TB, diabetic retinopathy screenings. *CodeBlue*. https://codeblue.galencentre.org/2025/03/moh-exploring-ai-for-cancer-tb-diabetic-retinopathy-screenings
10. Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*(4), 281–302. https://doi.org/10.1037/h0040957
11. DeVellis, R. F. (2017). *Scale development: Theory and applications* (4th ed.). SAGE Publications.
12. Dwivedi, Y. K., Rana, N. P., Jeyaraj, A., Clement, M., & Williams, M. D. (2019). Re-examining the unified theory of acceptance and use of technology (UTAUT): Towards a revised theoretical model. *Information Systems Frontiers, 21*(3), 719–734. https://doi.org/10.1007/s10796-017-9774-y
13. Economic Planning Unit. (2021). *Twelfth Malaysia Plan 2021–2025: A prosperous, inclusive, sustainable Malaysia*. Prime Minister's Department.
14. 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
15. Gefen, D., Karahanna, E., & Straub, D. W. (2003). Trust and TAM in online shopping: An integrated model. *MIS Quarterly, 27*(1), 51–90. https://doi.org/10.2307/30036519
16. Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2021). *A primer on partial least squares structural equation modeling (PLS-SEM)* (3rd ed.). SAGE Publications.
17. Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review, 31*(1), 2–24. https://doi.org/10.1108/EBR-11-2018-0203
18. 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
19. Henseler, J., Ringle, C. M., & Sarstedt, M. (2016). Testing measurement invariance of composites using partial least squares. *International Marketing Review, 33*(3), 405–431. https://doi.org/10.1108/IMR-09-2014-0304
20. Inland Revenue Board of Malaysia. (2025). *Tax audit framework on income tax and employer*. https://www.hasil.gov.my/media/wgpombsp/20250315_tax-audit-framework-on-income-tax-and-employer.pdf
21. Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research, 35*(6), 382–386. https://doi.org/10.1097/00006199-198611000-00017
22. Mahusin, N., Sallehudin, H., & Satar, N. S. M. (2024). Malaysia public sector challenges of implementation of artificial intelligence (AI). *IEEE Access, 12*, 121035–121051. https://doi.org/10.1109/ACCESS.2024.3448311
23. MAMPU. (2021). *Public sector digitalisation strategic plan 2021–2025* [Pelan Strategik Pendigitalan Sektor Awam 2021–2025]. Prime Minister's Department.
24. McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: An integrative typology. *Information Systems Research, 13*(3), 334–359. https://doi.org/10.1287/isre.13.3.334.81

25. Medrano-Sánchez, E. J., & Jara, J. (2026). Artificial intelligence agents as advanced decision support systems in public decision-making: Evidence from Peru. *Frontiers in Artificial Intelligence, 9*, 1805539. https://doi.org/10.3389/frai.2026.1805539
26. Ministry of Science, Technology and Innovation. (2021). *National artificial intelligence roadmap 2021–2025*. Government of Malaysia.
27. Moussawi, S., & Koufaris, M. (2019). Perceived intelligence and perceived anthropomorphism of personal intelligent agents: Scale development and validation. In *Proceedings of the 52nd Hawaii International Conference on System Sciences*. https://doi.org/10.24251/HICSS.2019.015
28. Moussawi, S., Koufaris, M., & Benbunan-Fich, R. (2021). How perceptions of intelligence and anthropomorphism affect adoption of personal intelligent agents. *Electronic Markets, 31*(2), 343–364. https://doi.org/10.1007/s12525-020-00411-w
29. Moussawi, S., Koufaris, M., & Benbunan-Fich, R. (2023). The role of user perceptions of intelligence, anthropomorphism, and self-extension on continuance of use of personal intelligent agents. *European Journal of Information Systems, 32*(3), 601–622. https://doi.org/10.1080/0960085X.2021.2018365
30. Musa, H., Azmi, F. R., Khalid, F. A., & Yunus, A. R. (2016). Analysing the adoption of Malaysian Standards (MS) on SME performances: The mediating effects of innovation adoption factors. In *Proceedings of the 2016 International Conference on Industrial Engineering and Operations Management*, Kuala Lumpur, Malaysia. IEOM Society. http://ieomsociety.org/ieom_2016/pdfs/258.pdf
31. Nawafleh, S., & Fares, A. M. S. (2024). UTAUT and determinant factors for adopting e-government in Jordan using a structural equation modelling approach. *Electronic Government, 20*(1), 20–46. https://doi.org/10.1504/EG.2024.135323
32. Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.
33. Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors, 52*(3), 381–410. https://doi.org/10.1177/0018720810376055
34. Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health, 29*(5), 489–497. https://doi.org/10.1002/nur.20147
35. Rai, A., Constantinides, P., & Sarker, S. (2019). Editor's comments: Next-generation digital platforms: Toward human–AI hybrids. *MIS Quarterly, 43(1), iii–ix. https://misq.umn.edu/pub/skin/frontend/default/misq/pdf/V43I1/EdComments_V43_I1.pdf*
36. Sapian, M. S. M., Musa, H., Rashid, N., Azmi, F. R., Ibrahim, M. S., & Low, Y. Y. (2026a). AI-DSS adoption readiness across health, education, and finance ministries: A cross-sectoral policy review. *European Journal of Prosthodontics and Restorative Dentistry, 34*(14s), 52–61. https://doi.org/10.1922/ejprd.v34i14s.2214
37. Sapian, M. S. M., Musa, H., Rashid, N., Ibrahim, M. S., & Kamarudin, S. M. (2026b). Extending the UTAUT model for artificial intelligence adoption: A systematic review of the role of perceived intelligence and trust. *International Journal of Research and Innovation in Social Science, 10*(6), 17711–17722. https://doi.org/10.47772/IJRISS.2026.100601221
38. Sapian, M. S. M., Musa, H., Rashid, N., Ibrahim, M. S., & Kamarudin, S. M. (2026c). Perceived intelligence and AI-DSS adoption in the public sector: A systematic review of UTAUT extensions. In Z. Mustafa & E. Mohamad (Eds.), *Proceedings of International Innovative Research and Innovative Dialogue (iIRID 2026): Fostering industry–academia innovation: Sustainable design, technopreneurship, and quality systems* (pp. 108–114). Penerbit UTeM Press. https://iirid.utem.edu.my/proceeding-2026/
39. Sapian, M. S. M., Musa, H., Rashid, N., Ibrahim, M. S., Krishnan, I. A., Low, Y. Y., & Kamarudin, S. M. (2026d). Determinants and challenges of artificial intelligence adoption in public sector governance: A systematic review. *International Journal of Research and Innovation in Social Science, 10*(8), 10029–10042. https://doi.org/10.47772/IJRISS.2026.100800676
40. Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *International Journal of Human-Computer Studies, 146*, 102551. https://doi.org/10.1016/j.ijhcs.2020.102551
41. Tamilmani, K., Rana, N. P., & Dwivedi, Y. K. (2021). The extended unified theory of acceptance and use of technology (UTAUT2): A systematic literature review and theory evaluation. *International Journal of Information Management, 57*, 102269. https://doi.org/10.1016/j.ijinfomgt.2020.102269

42. 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
43. United Nations Department of Economic and Social Affairs. (2022). *E-Government survey 2022: The future of digital government*. United Nations. https://publicadministration.un.org/egovkb/en-us/Reports/UN-E-Government-Survey-2022
44. Venkatesh, V., & Morris, M. G. (2000). Why don't men ever stop to ask for directions? Gender, social influence, and their role in technology acceptance and usage behavior. *MIS Quarterly, 24*(1), 115–139. https://doi.org/10.2307/3250981
45. Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly, 27*(3), 425–478. https://doi.org/10.2307/30036540
46. World Bank. (2021). *World development report 2021: Data for better lives*. World Bank. https://doi.org/10.1596/978-1-4648-1600-0