ISSN 2669-2481 / eISSN 2669-249X

VILNIUS TECH
Vilnius Gediminas Technical University

BUSINESS, MANAGEMENT AND ECONOMICS ENGINEERING

2026
Volume 24
Issue 2
Pages 387–406

https://doi.org/10.3846/bmee.2026.27100

VALIDATING AN AHP-BASED DECISION-SUPPORT FRAMEWORK FOR AI ADOPTION IN PUBLIC ADMINISTRATION

Primož PEVCIN✉, Katja DEBELAK, Rok HRŽICA

Faculty of Public Administration, University of Ljubljana, Ljubljana, Slovenia

Article History:
- received 17 May 2026
- accepted 10 July 2026

Abstract. *Purpose* – This study validates a multi-criteria decision-support framework for assessing organisational readiness for Artificial Intelligence (AI) adoption in public administration, with a specific focus on municipal administration. Validation ensures that the results are conceptually sound, credible, reliable, and robust.

Research methodology – The framework integrates the Technology-Organisation-Environment (TOE) and Unified Theory of Acceptance and Use of Technology (UTAUT) approaches into a hierarchical readiness structure, operationalised through the Analytical Hierarchy Process (AHP). Expert judgments were aggregated using pairwise comparisons to derive criteria' weights and support group decision-making. Validation combined expert assessment of conceptual coherence, completeness, clarity, interpretability, scale suitability, and practical utility with robustness testing across different Multi-Criteria Decision-Making methods (MCDM).

Findings – Voluntariness of use, behavioural intention to use AI, social influence, innovation and readiness for change, skills and expertise, data, and leadership represent central AI-readiness conditions. Expert validation supports the relevance and usability of the proposed hierarchy for AI-readiness assessment. Rankings across alternative MCDM methods showed strong convergence, indicating stable assessment outcomes despite differences in aggregation logic.

Research limitations – The framework was validated using three empirical organisational cases in public administration, along with five synthetically constructed readiness profiles for methodological testing, which may limit the generalisability of the findings.

Practical implications – The framework helps administrations diagnose AI-readiness gaps, benchmark organisational capabilities, and prioritise improvement measures related to AI adoption.

Originality/Value – The study contributes to a validated AI-readiness assessment framework that integrates structured expert judgment with multi-method robustness testing in the context of public administration.

Keywords: expert validation, MCDM methods, decision-support framework, AHP, AI adoption.

JEL Classification: C88, D81, O33.

✉Corresponding author. E-mail: *primoz.pevcin@fu.uni-lj.si*

1. Introduction

AI adoption in public administration is not only a technological decision, but also a multidimensional readiness problem involving organisational capacity, data maturity, leadership support, employee acceptance, regulatory constraints, and implementation conditions. Public administrations are expected to adopt digital and AI-based solutions while maintaining accountability, transparency, procedural reliability, and public value. These requirements make AI adoption more complex than conventional technology implementation. In municipal administrations, the challenge is especially pronounced because adoption decisions are often

made under conditions of limited resources, heterogeneous data infrastructures, uneven digital capabilities, and uncertain regulatory expectations (Mergel et al., 2019; Neumann et al., 2024; Wirtz et al., 2019).

Despite growing interest in artificial intelligence adoption in public administration, many readiness assessments remain fragmented across technological, organizational, institutional, and individual acceptance dimensions. This fragmentation makes it difficult for public managers to identify capability gaps, compare organisational readiness, and prioritise interventions. A useful readiness assessment must therefore do more than just list factors associated with AI adoption. It must organise these factors into a coherent structure, support transparent expert judgment, and produce diagnostic results that are interpretable for decision-makers (Jöhnk et al., 2021; Mikalef & Gupta, 2021).

This study addresses this need by validating an AHP-based decision-support framework for assessing AI readiness in public administration. The framework defines AI readiness through an integrated TOE and UTAUT hierarchy. TOE provides a basis for assessing technological, organisational, and environmental readiness conditions, while UTAUT captures individual acceptance and use-related determinants (Tornatzky et al., 1990; Venkatesh et al., 2003). AHP structures and aggregates expert judgments into criteria weights, while alternative MCDM methods test whether the framework produces stable rankings across different aggregation logics (Forman & Gass, 2001; Saaty, 2008). The methodological problem addressed in this paper is thus also how to validate a decision-support framework that structures expert judgments in a transparent, theoretically defensible, and practically usable manner. This distinction is important because expert-based readiness models cannot be considered valid simply because they generate numerical weights or rankings. Their validity depends on whether the decision hierarchy adequately represents the construct of AI readiness, whether experts find the criteria meaningful and interpretable, and whether the results remain stable when different decision-analytic procedures are applied (Cronbach & Meehl, 1955; Hemming et al., 2018; O'Hagan, 2019; Strauss & Smith, 2009).

Through systematic elicitation, hierarchical decomposition, and transparent weighting, expert judgments become explicit, comparable, and suitable for structured aggregation in uncertain public administration decision-making environments. This is especially important because relevant criteria in such contexts are often qualitative, interdependent, and context-sensitive, which reduces the utility of standardised indicators and increases the value of informed expert assessment (Morgan, 2014; Petkov et al., 2007). However, expert knowledge is not self-validating. Its value depends on disciplined procedures that ensure transparency, consistency, and defensibility (Hemming et al., 2018; O'Hagan, 2019; Salomon & Gomes, 2024).

Beyond expert assessment, an additional methodological challenge concerns the robustness of the decision-support framework used to aggregate these judgments. Despite extensive development of new MCDM techniques and hybrid approaches, validation efforts have typically focused on efficiency comparisons, consistency analysis, or illustration through real-world cases (Wu & Tiao, 2018). However, an important methodological challenge remains in determining which MCDM method is most appropriate for a given decision context, especially when reliable preference elicitation from decision-makers is difficult. Because MCDM

outputs can significantly influence real-world choices, it is important to examine whether different methods produce divergent rankings, as such differences may affect the final decision outcome (Vassoney et al., 2021).

Thus, the validation focuses on two dimensions. First, the study examines whether the proposed TOE-UTAUT hierarchy provides a conceptually coherent, complete, clear, interpretable, and practically usable representation of AI readiness. Second, it examines whether the rankings generated by the framework remain stable across alternative MCDM aggregation methods. Accordingly, the study addresses the following research questions:

RQ1. To what extent do experts assess the proposed AHP-based TOE-UTAUT AI-readiness framework as conceptually coherent, complete, clear, interpretable, and practically usable for public administration?

RQ2. To what extent are the assessment results generated by the decision-support framework stable across alternative MCDM aggregation methods?

The study makes three main contributions. First, it advances theory by operationalising the TOE-UTAUT model as a structured AI-readiness hierarchy for public administration. Second, it contributes methodologically by integrating AHP-based expert elicitation with cross-method robustness testing to validate a TOE-UTAUT AI-readiness decision-support framework. Third, it offers practical value by providing municipalities with a diagnostic tool to identify AI-readiness gaps and prioritise interventions.

Validating such a tool in this context is essential because a decision-support framework for AI readiness must do more than generate results; it must ensure that those results are conceptually sound, credible, and reliable for decision-making. Robustness testing across alternative MCDM methods further ensures that outcomes are stable and not dependent on a specific aggregation technique. This enhances trust in the framework and makes it suitable for real-world applications, where decision-makers require defensible, transparent, and methodologically sound assessments to allocate resources and prioritize AI initiatives. Although the tool is generally titled, it is important to emphasise that it is specifically tailored for municipal administrations, which represent only one segment of public administration. Different layers of public administration face distinct constraints, resource availability, capacities, and accountability requirements. Generic frameworks may fail to capture these differences, highlighting the need for a tailored and validated tool that reflects context-specific conditions and ensures that results are credible, interpretable, and stable.

2. Theoretical framework

Because the aim of this study is to validate a decision-support framework rather than develop new theory, this section first establishes the theoretical foundations of the AI-readiness construct and then derives the conceptual and methodological requirements that the framework and its validation must meet.

2.1. AI readiness in public administration

Theoretically, AI readiness should be a distinct construct, not a synonym for technology adoption or user acceptance. Readiness refers to the extent to which an organisation has

the necessary preconditions, i.e., technological, organisational, environmental, and individual ones, that must be in place before adoption can succeed. It is therefore antecedent to, rather than equivalent to, actual adoption (Jöhnk et al., 2021). This distinction is important in public administration, where formal decisions to adopt AI often come before the development of the organisational and human capabilities needed for implementation. Conceptualizing readiness as a multidimensional and sociotechnical construct also clarifies why single-factor or purely technological assessments are inadequate: readiness results from the interaction of data and infrastructure maturity, organisational capacity and leadership, regulatory and institutional conditions, and employee acceptance (Mikalef & Gupta, 2021; Neumann et al., 2024; Wirtz et al., 2019). This understanding provides the theoretical basis for operationalising readiness through an integrated, multilevel hierarchy rather than a flat list of factors.

AI readiness in public administration refers to the capacity of public organisations to adopt, implement, and use artificial intelligence in ways that are technically feasible, organizationally supported, legally compliant, and accepted by users. Unlike purely technological adoption problems, AI readiness in public administration involves administrative routines, accountability requirements, data governance, leadership commitment, employee competencies, interorganisational dependencies, and citizen-facing implications. This makes AI readiness a sociotechnical and managerial construct rather than a narrow measure of technological availability (Jöhnk et al., 2021; Mikalef & Gupta, 2021; Wirtz et al., 2019).

For municipal administrations in particular, readiness assessment is especially important because AI adoption decisions are often made with limited resources, fragmented data infrastructures, uneven digital capabilities, and uncertain regulatory expectations. They also operate under specific institutional constraints, including procedural accountability, legal compliance, political oversight, and the obligation to protect public value (Mergel et al., 2019; Neumann et al., 2024). A useful assessment framework must therefore integrate organisational, technological, environmental, and individual acceptance factors. It must also provide public managers with interpretable outputs that guide prioritisation, rather than only producing abstract scores.

2.2. Conceptual and methodological basis: TOE-UTAUT structured expert judgment

The conceptual structure of the proposed decision-support framework integrates TOE and UTAUT frameworks. TOE addresses the organisational and contextual factors influencing technology adoption, such as technological infrastructure, organisational capacity, and external pressures (Tornatzky et al., 1990). UTAUT complements this by focusing on individual acceptance and utilisation determinants, including performance expectancy, effort expectancy, social influence, facilitating conditions, behavioural intention, and voluntariness of use (Venkatesh et al., 2003). Integrating TOE and UTAUT is appropriate for AI-readiness assessment because AI adoption in public administration depends on both institutional capacity and individual acceptance. An administration may have technological infrastructure and formal leadership support, yet adoption may still fail if employees do not perceive AI as useful, if adequate skills and facilitating conditions are lacking, or if organisational routines discourage experimentation. Conversely, individual willingness to use AI is insufficient without data

quality, governance capacity, managerial support, and resources. The combined TOE-UTAUT framework thus reflects the multilevel nature of AI readiness. Besides, AI adoption in public organisations involves not only technological feasibility, but also administrative legitimacy, organisational capacity, employee acceptance, and alignment with public values. Prior research on AI in public administration shows that technological and organisational factors vary across adoption stages and that public organizations face specific challenges related to accountability, institutional constraints, and implementation capacity (Neumann et al., 2024; Wirtz et al., 2019).

Each readiness dimension is grounded in an established line of research. The technological dimension draws on innovation-adoption theory, which links adoption to a technology's relative advantage, compatibility, and complexity (Tornatzky et al., 1990). The organisational dimension reflects resource- and capability-based explanations of adoption, where leadership commitment, skills, and available resources shape absorptive capacity (Mikalef & Gupta, 2021). The environmental dimension addresses institutional and regulatory pressures specific to public administration, where legitimacy and compliance influence adoption as much as efficiency (Mergel et al., 2019; Wirtz et al., 2019). The individual dimension is grounded in acceptance theory, which explains use through performance expectancy, effort expectancy, social influence, and behavioural intention (Venkatesh et al., 2003). Specifying the theoretical origin of each dimension justifies its inclusion in the hierarchy and distinguishes the framework from an atheoretical checklist of adoption factors.

Expert judgment is central to AI-readiness assessment because many relevant criteria are qualitative, context-dependent, and difficult to capture with standardised indicators. In public-administration settings, experts can identify diagnostic cues related to organisational routines, leadership commitment, data maturity, implementation capacity, and acceptance barriers. However, expert judgment must be structured to avoid ambiguity, inconsistency, and implicit bias. Structured elicitation improves transparency and defensibility by making assumptions explicit and supporting systematic comparison across criteria (Hemming et al., 2018; Morgan, 2014; O'Hagan, 2019).

AHP is well suited for this purpose because it decomposes complex decision problems into hierarchical levels and uses pairwise comparisons to elicit expert preferences (Forman & Gass, 2001; Saaty, 2008). By requiring systematic comparisons, it transforms tacit professional knowledge into explicit and transparent priority structures, making expert judgments comparable and suitable for aggregation. Its consistency logic further strengthens methodological defensibility by enabling assessment of whether expert judgments form a coherent pattern (Salomon & Gomes, 2024). This is particularly valuable in public administration contexts characterized by multiple stakeholders, conflicting objectives, qualitative criteria, and uncertainty, as AHP both organizes expert knowledge into a structured framework and supports clear communication of how criteria weights are derived.

2.3. Validation logic and robustness of the framework

Validation of the decision-support framework extends beyond demonstrating its ability to produce rankings. It is conceived as the accumulation of evidence that the framework captures the construct of AI readiness in a theoretically sound, expert-interpretable, and

practically useful manner. This integrated approach combines construct and content validity with cognitive credibility, systems thinking, and methodological robustness to ensure coherence, relevance, and usability.

Construct validity concerns whether the framework adequately represents the theoretical construct it is intended to measure (Cronbach & Meehl, 1955; Strauss & Smith, 2009). In this context, the decision hierarchy must reflect the multidimensional nature of AI readiness rather than assembling unrelated indicators. Content validity concerns whether the selected criteria comprehensively cover the relevant domain of AI readiness in public administration (Carmines & Woods, 2005). Cognitive credibility concerns whether experts can meaningfully interpret the criteria and provide judgments that reflect their domain knowledge (Anders Ericsson & Towne, 2010; Feltovich et al., 2006). Systems thinking concerns whether the hierarchy captures the interdependence of technological, organisational, environmental, and individual factors embedded in public-administration systems (Petkov et al., 2007).

This validation logic directly informs the empirical design. Expert validation assesses the relevance, completeness, clarity, interpretability, scale suitability, and practical utility of the criteria. Cross-method comparison examines whether the rankings remain stable across different methods. Together, these procedures evaluate whether the proposed framework functions as a scientifically defensible decision-support tool rather than a one-time application of an MCDM technique. Namely, a decision-support framework is more credible when its outputs are not highly dependent on a single aggregation method. In MCDM research, robustness refers to the stability of rankings under different modelling assumptions, weighting procedures, normalisation strategies, or aggregation logics (Triantaphyllou, 2000; Velasquez & Hester, 2013). When several conceptually different methods produce similar rankings, this suggests that the decision-support framework is not overly sensitive to the selection of the methods. Different methods can generate different rankings because they rely on varying assumptions about compensation, normalisation, distance from ideal solutions, or rule-based qualitative classification (Sahoo & Goswami, 2023; Siksnelyte-Butkiene et al., 2020; Zlaugotne et al., 2020).

Therefore, comparing methods can serve as a validation strategy when the research objective is to determine whether a decision-support framework produces stable and interpretable outputs. In this study, WSM, WPM, WASPAS, DEX, and TOPSIS are used as robustness checks. Because all methods use the same AHP-derived weights, differences in rankings can be attributed to aggregation logic rather than to different weighting procedures. This design allows the robustness of the decision-support framework to be assessed independently of the weighting stage.

3. Methodology

3.1. Research design and decision-support framework

This study employs a validation-oriented MCDM design that integrates expert elicitation, AHP-based weighting, expert panel review, and robustness testing using several complementary aggregation techniques. The methodology aims not only to calculate criteria weights or rank alternatives, but also to assess whether the proposed AI-readiness framework serves as

a conceptually sound, interpretable, and practically usable decision-support tool for public administration.

Framework validation emphasizes the conceptual soundness of the hierarchical structure rather than relying solely on numerical priority values. Because AHP weights reflect contextual expert judgment, model validity is assessed through construct validity and content validity. Construct validity refers to the extent to which the hierarchy represents the theoretical construct of AI readiness (Cronbach & Meehl, 1955; Strauss & Smith, 2009). Content validity addresses whether the selected criteria comprehensively capture the relevant domain of AI readiness in public administration (Carmines & Woods, 2005).

Validation uses a mixed-method, multi-stage design that combines expert panel review with multi-method MCDM assessment. The expert panel review assesses the conceptual and practical adequacy of the framework, while the second validation stage tests whether the framework produces stable results across different aggregation procedures.

The framework is organized as a hierarchical AI-readiness model. The upper level contains the main readiness dimensions derived from the integrated TOE-UTAUT logic. The lower level contains operational criteria representing specific technological, organisational, environmental, and individual acceptance conditions. This hierarchical structure allows complex readiness conditions to be broken down into criteria that experts can compare, and organisations can assess.

The framework links theoretical representation with practical diagnosis, as indicated in Table 1. TOE provides the structure for assessing technological, organisational, and environmental readiness, while UTAUT adds individual acceptance and use-related determinants. As already noted, this combined structure is suitable for public administration because AI adoption depends on organizational capacity, data and technological infrastructure, leadership, external conditions, employee acceptance, and perceived usefulness.

Table 1. Validation logic of the AI-readiness decision-support framework (source: authors' presentation)

Validation dimension	Purpose	Operationalization in this study	Theoretical basis
Construct validity	To assess whether the framework represents the theoretical construct of AI readiness	The hierarchy operationalizes AI readiness through an integrated TOE-UTAUT structure that combines institutional capacity and individual acceptance factors	Cronbach and Meehl (1955), Strauss and Smith (2009)
Content validity	To assess whether the criteria comprehensively capture the relevant domain of AI readiness in public administration	Expert review evaluates whether the criteria are relevant, complete, and appropriate for municipal AI-readiness assessment	Carmines and Woods (2005)
Structured expert judgment	To elicit and aggregate expert knowledge in a transparent and systematic way	AHP pairwise comparisons are used to derive criteria weights, and expert matrices are aggregated into consolidated priority structures	Forman and Gass (2001), Saaty (2008)

End of Table 1

Validation dimension	Purpose	Operationalization in this study	Theoretical basis
Cognitive credibility	To assess whether experts can understand, interpret, and evaluate the hierarchy meaningfully	Experts assess conceptual clarity, statement clarity, scale suitability, and interpretability of results	Anders Ericsson and Towne (2010), Hemming et al. (2018)
Practical applicability	To assess whether the framework can support real public-sector decision-making	Experts assess the usability of the framework as a diagnostic tool for municipal AI readiness	Petkov et al. (2007)
Method-ological robustness	To assess whether results remain stable across different aggregation procedures	Rankings are compared across TOPSIS, WSM, WPM, WASPAS, and DEX	Triantaphyllou (2000), Velasquez and Hester (2013), Sahoo and Goswami (2023), Siksnelyte-Butkiene et al. (2020), Yannis et al. (2020), Zlaugotne et al. (2020)
Method selection adequacy	To justify the use of several complementary MCDM techniques rather than one method	Classification-oriented and aggregation-based techniques are combined to capture different decision logics	Cinelli et al. (2023), Triantaphyllou (2000)

Table 1 thus serves as the methodological link between the theoretical structure of the framework and the validation procedures described in the following sections. Together, these validation dimensions translate the abstract requirement of framework validity into concrete, assessable criteria and structure the empirical validation procedures described in the following sections.

3.2. Expert elicitation, validation and AHP weighting

AHP is used as a structured method for eliciting, structuring, and aggregating expert judgment (Forman & Gass, 2001; Saaty, 2008). Expert input was collected through two complementary procedures. First, 59 experts provided pairwise comparisons of criteria within the hierarchy. These comparisons determined the criteria weights for the subsequent MCDM analyses. Second, an expert validation panel assessed the conceptual and practical adequacy of the proposed framework using the validation dimensions summarised in Table 1.

Experts were selected based on their professional roles and experience in public administration, municipal management, digital transformation, AI adoption, and organisational decision-making. Compliance with these selection criteria was verified through their professional backgrounds and institutional roles. This ensured that pairwise comparisons were provided by participants capable of assessing the technological, organisational, environmental, and individual readiness conditions relevant to AI adoption in municipal administration.

Specifically, the second stage expert validation panel involved 25 experts with substantial experience in public-sector governance, digital transformation, and AI adoption. Participants were purposively selected based on their expertise in AI and public administration, with

representation from different types of public-sector institutions and governance levels. Expert status was verified through educational attainment, professional position, experience, and institutional affiliation. The panel combined policymakers, public managers, and academic researchers, thereby strengthening the credibility and diversity of perspectives reflected in the findings. Moreover, the approach followed standard prescriptions from the expert validation literature (Johansen & Fischer-Hübner, 2023), and validation panel structure is illustrated in Table 2.

Table 2. Expert validation panel selection criteria and participants' characteristics (source: authors' presentation)

Selection criterion	Participant characteristics
Experience with AI-related initiatives	Participants had professional experience with AI-related projects, initiatives or applications.
Familiarity with public administration processes	Participants possessed knowledge of public administration procedures and decision-making processes.
Relevant professional functions	Participants were involved in managerial, policymaking, analytical or expert functions.
Professional roles	Participants were macro-level policy makers (7), micro-level policy makers (4), or academic (14).
Institutional representation	Participants represented central and local government, public agencies and academia.
Professional experiences	Participants possessed following experience: from 5 to 10 years (6), from 10 to 20 years (9), more than 20 years (10).

AHP is used only to derive criteria weights through pairwise comparisons and is not employed as the final ranking procedure. This distinction is central to the validation design because it separates the elicitation and weighting of expert judgments from the subsequent robustness assessment of rankings. The AHP-based weights are then used as common inputs for several independent MCDM methods that assess organisations according to different aggregation logics.

Group decision-making was facilitated by aggregating individual pairwise comparison matrices into group comparison matrices. At each level of the hierarchy, the corresponding pairwise judgments from the participating experts were combined using the geometric mean. Each value in the group comparison matrix was calculated from the corresponding values in the individual expert matrices. The aggregated group matrices were then used to calculate priority vectors using the AHP eigenvalue method (Forman & Gass, 2001; Saaty, 2008). These priority vectors represented the local criteria weights and served as input for the subsequent robustness assessment. Consistency was assessed for the aggregated group matrices using the AHP Consistency Ratio (CR). All aggregated matrices met the accepted AHP consistency threshold of $CR < 0.10$. This procedure makes expert judgment explicit, comparable, and suitable for systematic evaluation. It also reduces reliance on a single evaluator by producing consensus-based priority structures from aggregated expert matrices.

The expert validation panel assessed the framework's relevance, conceptual clarity, interpretability, and practical applicability. These validation dimensions correspond to conceptual

coherence (Cronbach & Meehl, 1955; Strauss & Smith, 2009), clarity and scale suitability (Anders Ericsson & Towne, 2010; Hemming et al., 2018), as well as interpretability and practical applicability (Petkov et al., 2007). Expert confirmation of the framework's appropriateness, clarity, and usability provides evidence that the framework is meaningful for municipal AI-readiness assessment. This expert validation stage is important because expert-based MCDM frameworks cannot be considered valid solely because they generate numerical outputs. They must also be understandable, conceptually coherent, and practically useful in the decision context where they are applied.

3.3. Robustness assessment, data transformation, and ranking comparison

A second validation stage tests methodological robustness by applying several distinct MCDM methods to the same criteria structure and AHP-derived weighting scheme. Convergent results across methods indicate structural stability and suggest that the ranking outcomes are not overly dependent on a single aggregation procedure (Triantaphyllou, 2000; Velasquez & Hester, 2013). Method selection is based on the major MCDM reviews and reflects the need to compare methods that differ in assumptions, aggregation logic, and treatment of compensation among criteria (Sahoo & Goswami, 2023; Siksnelyte-Butkiene et al., 2020; Yannis et al., 2020; Zlaugotne et al., 2020).

Given the complexity of MCDM method selection, the study uses several complementary techniques rather than relying on a single method (Cinelli et al., 2023; Triantaphyllou, 2000). The applied methods are TOPSIS, WSM, WPM, WASPAS, and DEX. WSM provides an additive compensatory assessment. WPM uses multiplicative aggregation and is therefore more sensitive to low criterion values. WASPAS combines additive and multiplicative components. TOPSIS ranks alternatives according to their relative distance from ideal and anti-ideal solutions. DEX offers a qualitative, rule-based classification approach. In the validation design, TOE and UTAUT define the criteria structure, AHP-derived weights establish the priority structure, and the additional MCDM methods test whether the framework produces stable results under different decision logics. DEX is included to complement the quantitative methods by examining whether the same readiness framework remains interpretable when assessed through qualitative, rule-based classification rather than only numerical aggregation. Together, these methods provide complementary assessments based on different aggregation logics.

The assessment included three empirical organisational cases and five synthetically constructed profiles. The empirical cases were based on actual organisational assessments, while the synthetic profiles were designed to test the framework's behaviour under different readiness configurations. The synthetic cases were: 1) an ideal high-readiness profile with all criteria assessed as "Yes", 2) a low-readiness profile with all criteria assessed as "No", 3) a neutral profile with all criteria assessed as "Maybe", 4) asymmetric profiles with strong performance in some dimensions and weak performance in others, and 5) a heterogeneous mixed profile combining different criterion assessments. This configuration enabled comparison of MCDM methods across both extreme and heterogeneous scenarios while keeping the number of interpretable alternatives manageable. The objective was methodological validation, not statistical representativeness, using sufficiently diverse readiness profiles to reveal potential

differences in aggregation behaviour and ranking stability. In addition, the comparison across methods is not intended to identify a single superior MCDM method; it serves only as a validation strategy. As noted, if different aggregation methods produce similar rankings when applied to the same criteria structure and weights, this provides evidence that the framework is robust across different aggregation procedures.

For the quantitative methods, expert ratings were converted to numerical values. The original scale used "No," "Maybe," and "Yes," coded as 0, 0.5, and 1. For WPM, the scale was adjusted to 0.01, 0.505, and 1 to prevent zero values from eliminating alternatives through multiplication. Since WSM and WPM use different numerical logic, absolute scores were not directly compared. The comparison focused on rankings. Ranking convergence was assessed both descriptively and statistically. The study compared top-ranked, mid-ranked, and bottom-ranked alternatives across methods. Kendall's coefficient of concordance, Spearman's rank correlation, and Kendall's tau coefficient were used to assess the degree of agreement among methods. This allowed assessment of whether the decision-support framework produced stable rankings across different aggregation logics.

4. Results

4.1. Expert validation and substantive readiness priorities

Experts assessed the framework as suitable for diagnosing municipal AI readiness. Their assessment supported the relevance of the criteria, the logical organisation of the hierarchy, the clarity of the statements, the suitability of the response scale, the interpretability of results, and the practical value of the framework. These findings provide evidence of content and cognitive validity for the proposed decision-support tool. Experts confirmed that the framework captures not only technological readiness, legal compliance, data protection, and technical security, but also organisational practices and work routines that may hinder AI implementation. Table 3 summarises the expert validation evidence and its implications for the framework.

AI readiness in municipal public administration depends on both systemic and individual-level conditions. Among the evaluated criteria, voluntariness of use, behavioural intention to use AI, social influence, innovation and readiness for change, skills and expertise, data, and leadership emerged as particularly influential readiness factors. The results also show that technological factors alone are insufficient to explain organisational AI readiness, as organisational and individual-level criteria received the highest overall weights. These findings suggest that the framework captures AI readiness as a sociotechnical construct rather than a purely technological capability. Experts found the framework especially effective for diagnosing municipal readiness for AI adoption. Beyond commonly cited barriers such as legal compliance, data protection, and technical security, they emphasised that established organisational practices and entrenched work routines may also hinder AI implementation. This highlights the importance of integrating organisational and individual acceptance factors into the readiness framework.

Table 3. Expert validation evidence (source: authors' presentation)

Validation criterion	Expert assessment	Implication for the framework
Relevance of criteria	Experts confirmed that the criteria reflect core conditions for AI readiness in municipal administration.	Supports content validity by showing that the selected criteria are appropriate for the assessment domain.
Completeness of structure	Experts confirmed that the framework includes the main technological, organisational, environmental, and individual-level dimensions of AI readiness.	Supports conceptual coverage and reduces the risk that important readiness dimensions are omitted.
Logical organisation of hierarchy	Experts assessed the hierarchy as logically structured and suitable for decomposing AI readiness into assessable criteria.	Supports construct validity by showing that the model structure corresponds to the multidimensional nature of AI readiness.
Clarity of statements	Experts confirmed that the criteria and assessment statements are understandable and sufficiently precise.	Supports cognitive credibility and indicates that experts can interpret the framework consistently.
Scale suitability	Experts assessed the "No," "Maybe," and "Yes" response scale as usable for evaluating readiness conditions.	Supports measurement feasibility and allows qualitative judgments to be translated into MCDM inputs.
Interpretability of results	Experts confirmed that the results can be meaningfully interpreted by practitioners and decision-makers.	Supports the diagnostic value of the framework for municipal AI-readiness assessment.
Practical utility	Experts assessed the framework as useful for identifying readiness gaps and supporting prioritisation of interventions.	Supports practical applicability as a decision-support tool for public administration.
Overall appropriateness	Experts confirmed the model's appropriateness, clarity, and usability for assessing AI readiness.	Provides combined validation evidence for the credibility and usability of the framework.

4.2. Ranking results and method-specific differences across MCDM methods

Table 4 presents the rankings obtained using different MCDM methods. Because all methods used the same AHP-derived criteria weights, differences in rankings reflect the aggregation logic of each method rather than differences in weighting.

Table 4. Final rankings of alternatives for five methods (source: authors' calculations)

Organisation	Ranking				
	WSM	WPM	WASPAS (λ = 0.5)	DEX	TOPSIS
Organisation 1	2.	2.	2.	1.–2.	2.
Organisation 2	3.	4.	4.	3.	3
Organisation 3	6.	6.	6.	6.	6.
Synthetic organisation 1	1.	1.	1.	1.–2.	1.
Synthetic organisation 2	8.	8.	8.	7.–8.	8.
Synthetic organisation 3	5.	3.	3.	4.–5.	5.
Synthetic organisation 4	7.	7.	7.	7.–8.	7.
Synthetic organisation 5	4.	5.	5.	4.–5.	4.

As expected, the uniformly high- and low-readiness synthetic profiles remain relatively stable across methods because consistently favourable or unfavourable criterion assessments tend to yield similar outcomes regardless of aggregation logic. Greater variation occurs among the mid-ranked alternatives, particularly for heterogeneous and moderate profiles, where compensatory and multiplicative aggregation approaches handle uneven criterion distributions differently. However, the observed ranking shifts are mainly limited to adjacent middle positions, while the overall ranking structure remains relatively stable across methods. Synthetic organisation 1 consistently ranks first across all quantitative methods, while Synthetic organisation 2 remains in the lowest position. Organisation 1 also maintains a consistently high ranking, whereas Organisation 3 stays in the lower part of the ranking across all methods. The main differences occur among Organisation 2, Synthetic organisation 3, and Synthetic organisation 5, whose relative positions vary slightly depending on the aggregation logic used.

Differences among methods appear mainly in the middle of the ranking, particularly between additive and multiplicative aggregation methods. The WPM approach is more sensitive to zero values because the overall score is calculated by multiplication. Consequently, alternatives with zero values for certain criteria receive lower final scores, as seen for Organisation 2 and Synthetic organisation 5 compared to WSM.

In contrast, Synthetic organisation 3, whose criteria values are uniformly moderate (e.g., around 0.5), performs relatively better using WPM and WASPAS. This behaviour reflects the fundamental difference between additive and multiplicative aggregation. In additive models such as WSM, strong performance on some criteria can compensate for weaker performance on others, whereas multiplicative models penalise low values more strongly because they directly reduce the overall product. As noted by Nasyuha et al. (2025) this characteristic makes WPM more sensitive to uneven performance profiles, while WSM tends to be more forgiving when high scores offset lower ones.

The WASPAS results closely follow the patterns observed in WSM and WPM, as the method combines additive and multiplicative aggregation. With λ set to 0.5, both components contribute equally to the final score, resulting in rankings that largely mirror those of the two underlying methods.

Sensitivity analysis shows that the relative positions of some mid-ranked alternatives depend on the weight assigned to the additive component. Organisation 2 improves its ranking as λ approaches the additive end of the scale ($\lambda \geq 0.9$), indicating a more favourable assessment under additive aggregation. In contrast, Synthetic organisation 5 achieves a higher position when $\lambda = 1$, that is, when the assessment corresponds to the pure WSM approach. This pattern suggests that multiplicative aggregation slightly penalises this alternative, while additive aggregation reduces the influence of lower criterion values. The underlying criterion-level assessments for the three alternatives whose positions vary across methods are presented in Table 5.

Table 5 presents the detailed criterion evaluation for the alternatives ranked third to fifth. A closer inspection of these alternatives further illustrates the different logics of aggregation. Alternatives with more balanced criteria values tend to perform better under WPM, as shown by Synthetic organisation 3, whose values across all criteria are consistently assessed

Table 5. Evaluation for the alternatives differently ranked (source: authors' calculations)

Basic criteria	Organisation 2	Synthetic organisation 3	Synthetic organisation 5
1.1. Perceived direct benefits of AI	Maybe	Maybe	Yes
1.2. Effort expectancy of AI	No	Maybe	Maybe
1.3. Data	No	Maybe	Yes
1.4. Technology	Maybe	Maybe	Maybe
2.1. Leadership	Maybe	Maybe	No
2.2. Skills and expertise	Maybe	Maybe	Yes
2.3. Innovation and readiness for change	Maybe	Maybe	Maybe
2.4. Costs and resources	No	Maybe	No
2.5. Facilitating conditions	Maybe	Maybe	Yes
3.1. Social influence	Maybe	Maybe	Maybe
3.2. State and citizens	Yes	Maybe	Yes
4.1. Behavioural intention to use AI	Maybe	Maybe	Maybe
4.2. Voluntariness of use	Yes	Maybe	No
Final score WSM	0.5417	0.5	0.5190
Final score WPM	0.3297	0.505	0.2174
Final score WASPAS (λ = 0.5)	0.4357	0.5025	0.3682
Final score DEX	High	Medium	Medium
Final score TOPSIS	0.5178	0.5	0.5027

as "Maybe". This results in a uniform performance profile and a relatively stronger position under WPM and WASPAS than under WSM. In contrast, alternatives with one or more "No" assessments experience a stronger negative effect under multiplicative aggregation. This is evident for Organisation 2 and Synthetic organisation 5, both of which include several "No" assessments and are therefore ranked slightly lower under WPM than under WSM.

DEX produces partially tied rankings because it uses rule-based classification instead of continuous numerical aggregation. Since DEX assigns alternatives to qualitative classes rather than calculating continuous numerical scores, direct score comparison with quantitative methods is not meaningful; only ranking and class correspondence can be interpreted. Alternatives are assigned to qualitative classes, and numerical differences within a class do not affect the final ranking. Consequently, alternatives within the same class receive identical final rankings, providing less granularity than continuous-value methods such as WSM, WPM, or WASPAS. If the additive WSM scores were grouped into five broader classification classes, similar groupings would appear, as the numerical differences among several mid-ranked alternatives are relatively small.

The TOPSIS results largely confirm the ordering obtained through aggregation-based methods, further supporting the stability of the assessment outcomes. Minor variations in the middle ranks are primarily due to the multiplicative logic of the WPM and do not substantially alter the overall ranking structure.

One methodological challenge with the WPM approach is that the choice of measurement scale can influence the final values. In multiplicative models, very low specific criterion' scores

can have a disproportionately strong negative effect on overall performance. Therefore, direct numerical comparisons between WSM and WPM are not meaningful, as WSM uses additive aggregation, which is less sensitive to individual low values, while WPM amplifies these differences. Because the two techniques operate on inherently different numerical scales (e.g., 0–1 vs. 0.01–1), only the ranking, rather than the absolute scores, can be validly compared. To address this issue, the performance scale was adjusted, from 0.0, 0.5, and 1, to 0.01, 0.505, and 1. This prevents extremely low values from reducing the overall score to zero and allows for more reasonable differentiation among alternatives. Even after the scale adjustment, the ranking remains stable, reinforcing the consistency of the results.

A conceptual difference between the methods is also evident. WSM and TOPSIS allow compensation across criteria, meaning weaker performance on one criterion can be offset by stronger performance on another. WSM is widely used because of its computational simplicity and efficiency, although it may be limited when complex relationships among criteria exist or when there are large disparities in criterion importance (Wiangkham & Vongvit, 2025). In contrast, WPM allows only limited compensation due to its multiplicative nature. As a result, a single very low criterion score can disproportionately reduce overall assessment and overshadow strong performance elsewhere. This characteristic explains why WPM behaves differently from the other techniques and why its absolute values should not be directly compared with WSM or TOPSIS scores. WASPAS is a hybrid approach that combines both additive and multiplicative aggregation, balancing compensatory and non-compensatory effects depending on the value of λ. DEX, by contrast, uses a rule-based qualitative assessment logic in which alternatives are classified into predefined categories rather than aggregated through numerical operations. Because of this categorical structure, DEX produces less granular results but offers an interpretable, rule-based perspective on the assessment.

4.3. Concordance among methods

Nevertheless, the comparison of rankings obtained from the five applied MCDM methods shows a remarkably high level of concordance. Kendall's coefficient of concordance indicates almost perfect overall agreement among the methods ($W = 0.9655$, $p < 0.001$), confirming that the observed consistency is highly unlikely to be due to chance. Pairwise analyses further support this: both Spearman's rank correlations ($\rho = 0.9286$–1.000) and Kendall's τ coefficients ($\tau = 0.8571$–1.000) show consistently strong and statistically significant associations across all method pairs. The classical weighting-based methods (WSM, WPM, WASPAS) and TOPSIS exhibit nearly identical ranking behaviour, while the DEX method, although slightly less aligned, still correlates very strongly with the others. Overall, the results indicate that all five MCDM methods produce highly consistent rankings, suggesting strong methodological consistency across the applied methods.

Thus, despite methodological differences and the specific characteristics of each technique, the rankings produced by WSM, WPM, WASPAS, DEX, and TOPSIS remain largely consistent. The highest- and lowest-ranked alternatives are stable across all methods, with only minor variations among the mid-ranked alternatives. This convergence across conceptually different decision approaches strengthens confidence in the assessment results and indicates that the observed ranking structure is robust and not strongly dependent on the decision method.

5. Discussion

The findings should be interpreted primarily as validation of the conceptual decision-support framework, not as confirmation of any specific weighting algorithm. Because all methods use the same AHP-derived weights, any differences in rankings result from differences in aggregation logic rather than weighting procedures. This design allows the robustness of the framework to be assessed independently of the weighting procedure. When several MCDM methods produce largely consistent results, this consistency provides strong evidence of methodological robustness. This interpretation directly addresses RQ1 by showing that the proposed TOE-UTAUT AI-readiness framework should be evaluated as a structured representation of AI readiness, rather than solely as a numerical output of AHP. The validation evidence therefore concerns the coherence of the hierarchy, the suitability of the criteria, and the framework's ability to organize expert judgment in a transparent, interpretable, and usable way.

In MCDM research, robustness refers to the stability of rankings despite variations in modelling assumptions, weighting schemes, or aggregation rules. Convergence across methods increases confidence in assessment results and indicates a methodologically robust decision-support framework (Avramova et al., 2025; Bączkiewicz et al., 2021). Because the proliferation of MCDM methods often leads to divergent outcomes for the same problem, many studies recommend comparing multiple techniques within the application context to support method selection and ensure credible rankings. As noted by Hien et al. (2025), algorithmic differences frequently produce inconsistent results; therefore, high agreement across methods signals methodological consistency and ranking stability. This directly addresses RQ2, as the stability of rankings across different MCDM methods indicates that the assessment results are relatively insensitive to a specific aggregation procedure.

Because all methods used identical AHP-derived weights, any ranking differences result solely from differences in aggregation logic rather than weighting procedures. This separation indicates that the framework functions coherently across analytical contexts, and that assessment results are not artifacts of a single aggregation method. Cognitive credibility further supports the framework, as experts found the hierarchy intuitive, interpretable, and aligned with their understanding of AI adoption challenges. When different MCDM methods produce similar top- and bottom-ranked alternatives, it indicates that the decision problem is structurally stable. Dominance relationships among alternatives are strong enough to persist despite differences in the logic of aggregation. This stability typically occurs when alternatives differ sufficiently in performance (Paradowski et al., 2025). Literature on MCDM robustness also emphasizes the importance of weight sensitivity. Rankings that remain stable under varying modelling conditions are generally considered more reliable. Fengmin et al. (2025) note that some methods, especially nonlinear ones, demonstrate greater robustness because they are less sensitive to weight thresholds. The observed stability reflects the structural robustness of the alternatives rather than effects driven by the weights. This finding is important because it separates the validation of the framework from the validation of any individual method. The results suggest that the framework captures sufficiently meaningful differences among alternatives to produce stable rankings even when the aggregation logic changes.

Convergence across methodologically diverse MCDM techniques suggests that the underlying framework (criteria selection, evaluation scales, and data quality) is coherent and well designed. Malefaki et al. (2025) demonstrate that the choice of MCDM method and normalisation strategy can meaningfully influence sustainability assessments. Therefore, convergence despite methodological variation indicates that the framework is coherent, internally consistent, and not overly dependent on specific methodological choices. To sum up, the high degree of agreement across MCDM methods shows that the alternatives differ enough to produce stable rankings regardless of the aggregation procedure. This indicates low methodological sensitivity and confirms that the framework is conceptually sound, coherent, and aligned with established notions of weight insensitivity, rank stability, and methodological triangulation. Therefore, the final rankings can be considered robust, reliable, and relatively insensitive to the choice of aggregation method within the set of methods tested.

6. Conclusions

The study validates a theoretically grounded decision hierarchy for assessing AI readiness in public administration and demonstrates how structured expert elicitation, combined with multiple decision-analytic methods, can evaluate the conceptual adequacy and robustness of complex readiness assessment frameworks. Conceptual validity is supported by the integrated TOE-UTAUT hierarchy, which effectively captures the multilevel determinants of AI readiness. The prominence of organisational and individual-level factors, such as social influence and voluntariness of use, confirms that the combined framework reflects the sociotechnical conditions shaping municipal AI adoption.

Content validity is reinforced by expert assessments indicating that the criteria meaningfully represent key readiness dimensions. Experts found the hierarchy conceptually coherent, complete, clear, interpretable, and practically usable. Experts consistently identified organisational culture, top management support, perceived usefulness, and data capabilities as central determinants, noting that entrenched organisational routines pose major barriers to AI adoption. The framework also demonstrated methodological robustness. Rankings produced by WSM, WPM, WASPAS, DEX, and TOPSIS were highly concordant despite differences in aggregation logic, indicating that the assessment results are relatively stable across the applied decision-analytic methods. This supports the robustness and practical usability of the proposed AI readiness assessment framework.

Several limitations should be acknowledged. The framework was validated in the context of municipal administration using only three empirical organisational cases, which may restrict the generalisability of the findings. In addition, part of the robustness assessment relied on synthetic readiness profiles designed for methodological testing rather than real organisational observations. The framework also depends on expert judgments and qualitative assessments, which may introduce contextual subjectivity despite the structured elicitation and aggregation procedures applied.

The findings provide practical value. Municipalities can use the framework as a structured diagnostic tool to identify strengths and gaps in AI readiness. The results emphasise the need for targeted interventions in organizational culture, leadership commitment, and data

maturity. Future research should expand the empirical base, integrate objective indicators, compare alternative weighting schemes, and test the framework across different public-sector contexts and over time.

Finally, the main implication is that the framework should be understood as a validated AI-readiness decision-support tool rather than a single-method ranking exercise. Its value lies in combining theoretically grounded criteria, structured expert elicitation, and cross-method robustness testing. This enhances its practical relevance for public administrations seeking transparent and methodologically defensible tools for diagnosing AI readiness and prioritising interventions.

Funding

The APC was funded by financial support from the internal UL FPA's project, "The impact of demographic change and artificial intelligence on the public sector labour market", funded by the Slovenian Research and Innovation Agency, Institutional Pillar of Financing.

AI acknowledgement

The language of this manuscript was edited with the assistance of InstaText (version 1.4.3) and Microsoft Copilot (GPT-5.5), which were used for grammar correction, language editing, sentence-level refinement, minor improvements to clarity and readability, and adaptation to British English. The tools were not used to generate research data, conduct analyses, interpret results, or draw the study's conclusions. The authors reviewed and edited the AI-assisted suggestions and remain responsible for the originality, validity, and integrity of the manuscript's content.

References

Anders Ericsson, K., & Towne, T. J. (2010). Expertise. *WIREs Cognitive Science*, *1*(3), 404–416. https://doi.org/10.1002/wcs.47

Avramova, T., Peneva, T., & Ivanov, A. (2025). Overview of existing multi-criteria decision-making (MCDM) methods used in industrial environments. *Technologies*, *13*(10), Article 444. https://doi.org/10.3390/technologies13100444

Bączkiewicz, A., Kizielewicz, B., Shekhovtsov, A., Wątróbski, J., & Sałabun, W. (2021). Methodical aspects of MCDM based e-commerce recommender system. *Journal of Theoretical and Applied Electronic Commerce Research*, *16*(6), 2192–2229. https://doi.org/10.3390/jtaer16060122

Carmines, E. G., & Woods, J. A. (2005). Validity assessment. In K. Kempf-Leonard (Ed.), *Encyclopedia of social measurement* (pp. 933–937). Elsevier. https://doi.org/10.1016/B0-12-369398-5/00418-7

Cinelli, M., Kadziński, M., Miebs, G., Słowiński, R., Gonzalez, M., & Burgherr, P. (2023). *MCDA Methods Selection Software* (*MCDA MSS*) [Computer software]. Poznan University of Technology. https://mcda.cs.put.poznan.pl/index.php

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281–302. https://doi.org/10.1037/h0040957

Feltovich, P. J., Prietula, M. J., & Ericsson, K. A. (2006). Studies of expertise from psychological perspectives. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (1st ed., pp. 41–68). Cambridge University Press. https://doi.org/10.1017/CBO9780511816796.004

Fengmin, L., Dengfeng, W., Ying, X., Zihao, M., & Jing, C. (2025). Classification, selection of MCDM methods and robust decision-making in multidisciplinary design optimization of automotive structures. *Structural and Multidisciplinary Optimization*, *68*, Article 232. https://doi.org/10.1007/s00158-025-04150-4

Forman, E. H., & Gass, S. I. (2001). The analytic hierarchy process – an exposition. *Operations Research*, *49*(4), 469–486. https://doi.org/10.1287/opre.49.4.469.11231

Hemming, V., Burgman, M. A., Hanea, A. M., McBride, M. F., & Wintle, B. C. (2018). A practical guide to structured expert elicitation using the IDEA protocol. *Methods in Ecology and Evolution*, *9*(1), 169–180. https://doi.org/10.1111/2041-210X.12857

Hien, N. T. T., Quynh, P. H., & Minh, V. Q. (2025). A comparative analysis of multi-criteria decision-making methods. *Engineering*, *Technology* & *Applied Science Research*, *15*(5), 26369–26375. https://doi.org/10.48084/etasr.12782

InstaText. (n.d.). *InstaText* (Version 1.4.3) [AI-assisted writing and editing tool]. https://instatext.io/

Johansen, J., & Fischer-Hübner, S. (2023). Expert opinions as a method of validating ideas: Applied to making GDPR usable. In N. Gerber, A. Stöver, & K. Marky (Eds.), *Human factors in privacy research* (pp. 137–152). Springer. https://doi.org/10.1007/978-3-031-28643-8_7

Jöhnk, J., Weißert, M., & Wyrtki, K. (2021). Ready or not, AI comes – an interview study of organizational AI readiness factors. *Business* & *Information Systems Engineering*, *63*, 5–20. https://doi.org/10.1007/s12599-020-00676-7

Malefaki, S., Markatos, D., Filippatos, A., & Pantelakis, S. (2025). A comparative analysis of multi-criteria decision-making methods and normalization techniques in holistic sustainability assessment for engineering applications. *Aerospace*, *12*(2), Article 100. https://doi.org/10.3390/aerospace12020100

Mergel, I., Edelmann, N., & Haug, N. (2019). Defining digital transformation: Results from expert interviews. *Government Information Quarterly*, *36*(4), Article 101385. https://doi.org/10.1016/j.giq.2019.06.002

Microsoft. (n.d.). *Microsoft Copilot* (GPT-5.5) [Large language model]. https://copilot.microsoft.com/

Mikalef, P., & Gupta, M. (2021). Artificial intelligence capability: Conceptualization, measurement calibration, and empirical study on its impact on organizational creativity and firm performance. *Information* & *Management*, *58*(3), Article 103434. https://doi.org/10.1016/j.im.2021.103434

Morgan, M. G. (2014). Use (and abuse) of expert elicitation in support of decision making for public policy. *Proceedings of the National Academy of Sciences*, *111*(20), 7176–7184. https://doi.org/10.1073/pnas.1319946111

Nasyuha, A. H., Tujantri, H., Veza, O., Nurarif, S., & Chung, M.-Y. (2025). Comparison of WSM and weight product methods with WSM-Score and vector approaches. *Sinkron*: *Jurnal Dan Penelitian Teknik Informatika*, *9*(2), 948–956. https://doi.org/10.33395/sinkron.v9i2.14817

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

O'Hagan, A. (2019). Expert knowledge elicitation: Subjective but scientific. *The American Statistician*, *73*(sup1), 69–81. https://doi.org/10.1080/00031305.2018.1518265

Paradowski, B., Wątróbski, J., & Sałabun, W. (2025). Novel coefficients for improved robustness in multi-criteria decision analysis. *Artificial Intelligence Review*, *58*, Article 298. https://doi.org/10.1007/s10462-025-11307-6

Petkov, D., Petkova, O., Andrew, T., & Nepal, T. (2007). Mixing multiple criteria decision making with soft systems thinking techniques for decision support in complex situations. *Decision Support Systems*, *43*(4), 1615–1629. https://doi.org/10.1016/j.dss.2006.03.006

Saaty, T. L. (2008). Decision making with the analytic hierarchy process. *International Journal of Services Sciences*, *1*(1), 83–98. https://doi.org/10.1504/IJSSCI.2008.017590

Sahoo, S. K., & Goswami, S. S. (2023). A comprehensive review of Multiple Criteria Decision-Making (MCDM) methods: Advancements, applications, and future directions. *Decision Making Advances*, *1*(1), 25–48. https://doi.org/10.31181/dma1120237

Salomon, V. A. P., & Gomes, L. F. A. M. (2024). Consistency improvement in the analytic hierarchy process. *Mathematics*, *12*(6), Article 828. https://doi.org/10.3390/math12060828

Siksnelyte-Butkiene, I., Zavadskas, E. K., & Streimikiene, D. (2020). Multi-Criteria Decision-Making (MCDM) for the assessment of renewable energy technologies in a household: A review. *Energies*, *13*(5), Article 1164. https://doi.org/10.3390/en13051164

Strauss, M. E., & Smith, G. T. (2009). Construct validity: Advances in theory and methodology. *Annual Review of Clinical Psychology*, *5*(1), 1–25. https://doi.org/10.1146/annurev.clinpsy.032408.153639

Tornatzky, L. G., Fleischer, M., & Chakrabarti, A. K. (1990). *The processes of technological innovation*. Lexington Books.

Triantaphyllou, E. (2000). Multi-criteria decision making methods. In *Multi-criteria decision making methods*: *A comparative study* (pp. 5–21). Springer. https://doi.org/10.1007/978-1-4757-3157-6_2

Vassoney, E., Mammoliti Mochet, A., Desiderio, E., Negro, G., Pilloni, M. G., & Comoglio, C. (2021). Comparing multi-criteria decision-making methods for the assessment of flow release scenarios from small hydropower plants in the Alpine area. *Frontiers in Environmental Science*, *9*, Article 635100. https://doi.org/10.3389/fenvs.2021.635100

Velasquez, M., & Hester, P. (2013). An analysis of multi-criteria decision making methods. *International Journal of Operations Research*, *10*(2), 56–66.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*: *Management Information Systems*, *27*(3), 425–478. https://doi.org/10.2307/30036540

Wiangkham, A., & Vongvit, R. (2025). Comparative analysis of multi-criteria decision making methods for prioritizing influential factors of ChatGPT adoption in higher education. *Expert Systems with Applications*, *287*, Article 128188. https://doi.org/10.1016/j.eswa.2025.128188

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

Wu, J.-Z., & Tiao, P.-J. (2018). A validation scheme for intelligent and effective multiple criteria decision-making. *Applied Soft Computing*, *68*, 866–872. https://doi.org/10.1016/j.asoc.2017.04.054

Yannis, G., Kopsacheili, A., Dragomanovits, A., & Petraki, V. (2020). State-of-the-art review on multi-criteria decision-making in the transport sector. *Journal of Traffic and Transportation Engineering* (*English Edition*), *7*(4), 413–431. https://doi.org/10.1016/j.jtte.2020.05.005

Zlaugotne, B., Zihare, L., Balode, L., Kalnbalkite, A., Khabdullin, A., & Blumberga, D. (2020). Multi-criteria decision analysis methods comparison. *Environmental and Climate Technologies*, *24*(1), 454–471. https://doi.org/10.2478/rtuect-2020-0028