

Article

Evaluation of Autonomous Vehicles Adoption Using the Technology Readiness Assessment Methodology: A Saudi Arabian Case Study

Asaeyl Alahmadi * , Salma Elhag and Maram Meccawy

Department of Information System, Faculty of Computing and Information Technology,
King Abdulaziz University, Jeddah 21589, Saudi Arabia; smomar@kau.edu.sa (S.E.)

* Correspondence: aalahmadi0396@stu.kau.edu.sa

Abstract

The transportation sector is a primary focus for Artificial Intelligence (AI) innovation because of ongoing technological progress. The sector faces growing challenges due to urbanization and congestion. Cities are considering electric and Autonomous Vehicles (AVs) to be promising future solutions because of their potential to enhance efficiency and reduce emissions, thereby helping us to achieve sustainability goals and drive the digital transformation. However, adopting these technologies requires evaluating their real-world deployment readiness. This research is motivated by the need to address this gap. The research develops and validates a smart framework through a technology readiness assessment methodology to evaluate AV readiness, combining rule-based reasoning with a weighted scoring and threshold-based classification model, along with Large Language Model (LLM) analysis. PESTEL and SWOT analyses were conducted to understand the current state of AV readiness in Saudi Arabia. The system was validated against a benchmark dataset derived from published sources of expert assessments across 20 countries. The system achieved a 97.1% agreement rate and a 0.44 Mean Absolute Error (MAE). The robustness of the weighted aggregation model was demonstrated through a Monte Carlo simulation, achieving a 92.1% stability rate when dealing with uncertain conditions. The framework proved its real-world value through a Saudi Arabian case study with an overall predicted TRL at 7–8, identifying social acceptance and regulation as a priority for development. These findings will advance AI adoption in the transport sector by providing a multidimensional framework to evaluate the readiness for deploying AVs.

Keywords: Autonomous Vehicles (AVs); Artificial Intelligence; smart mobility; Technology Readiness Level (TRL); Technology Readiness Assessment (TRA)



Academic Editors: Teng Zhou, Zhizhe Lin, Chunjie Cao and Youyi Song

Received: 1 July 2026

Revised: 30 July 2026

Accepted: 31 July 2026

Published: 31 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Smart cities have become a worldwide priority for urban development. The implementation of smart city technologies has created better urban management systems and improved citizen engagement [1]. Multiple countries have worked to achieve the complete benefits of smart city technologies. The combination of Artificial Intelligence (AI) with Internet of Things (IoT) and Big Data has enhanced sustainable urban development [2]. The establishment of smart cities creates business opportunities, which generate employment and economic development [1,3]. The development of smart cities in Saudi Arabia serves as a strategic initiative that supports Vision 2030 goals to improve life quality and support economic growth through the development of smart infrastructure, which includes building

energy networks, water management systems, and transportation networks. Smart cities achieve their targets through the implementation of several essential components, including smart transportation development. The implementation of smart transportation systems depends on sensors, electronics, computers, communication systems, and management strategies. These advancements result in improved traffic movement, reduced travel times, and lower fuel usage [4]. Autonomous Vehicles (AVs) represent a significant progress in intelligent transportation system [4]. AVs, known as self-driving cars, are vehicles that use advanced technologies to navigate roads independently without human intervention [4–6]. The vehicles operate through sensor systems, software, and AI technology to navigate roads, handle traffic situations, and perform driving tasks [4].

This research is motivated by the global push toward smart cities and smart mobility systems. AVs present promising solutions with their potential to reduce accidents, reduce transport emissions, and facilitate sustainable digital transformation. The global adoption of AVs is advancing, with distinct approaches in various regions, including China, the United States, the European Union, Germany, and Japan. The implementation of AVs throughout Saudi smart cities supports Vision 2030 goals because it delivers better life quality. This research addresses a specific gap in the adoption of AVs, as countries worldwide work to integrate this technology into their urban mobility and public transport sectors. The digital transformation of Saudi Arabia through Vision 2030 depends on the integration of AV technology to build a safer, more efficient, and sustainable future mobility system. However, this transition is hindered by multifaceted challenges, including infrastructure development, regulatory frameworks, and public acceptance. A comprehensive framework is needed to assess the readiness of AVs. A technology readiness assessment framework can play a pivotal role in bridging these gaps by offering a holistic approach that integrates infrastructure, technological, regulatory, data, and public acceptance dimensions. By providing this comprehensive analysis, it enables informed decision-making for successful AV adoption aligned with long-term sustainability goals. This framework is designed for urban planners and policymakers to support critical decisions regarding infrastructure and technology investment, regulatory planning, and strategic prioritization for AV integration. This research focuses on assessing readiness for highly automated vehicles within the domain of shared and public transport, excluding other sectors such as freight and delivery. This focus ensures a targeted and relevant analysis of the critical dimensions necessary for integrating AVs into future mobility systems. For this research, Saudi readiness for AVs is operationally defined as a profile composed of five key dimensions: infrastructure, technological, regulatory, data, and public acceptance. These five dimensions were systematically derived to ensure that all critical factors influencing AV deployment in the Saudi context were captured. The overall state of readiness is further characterized by a specific Technology Readiness Level (TRL) band, indicating the ecosystem's preparedness for a given phase of AV maturity, from testing to full-scale deployment. The smart assessment also provides a structured assessment process that moves readiness evaluation from expert-based subjective judgment to a more scalable, transparent, and consistent approach. The findings also aim to support the adoption of AVs in Saudi Arabia's transport sector, supporting the future goals of Vision 2030, which emphasizes technological innovation and sustainability in national development.

Existing readiness assessments tend to rely on a single data type, such as technical indicators, policy reviews, or market statistics, or apply black box scoring rules that obscure how qualitative evidence is translated into readiness ratings. This limits their usefulness for national policy design because decision-makers cannot trace recommendations back to the underlying evidence or assess the relative contributions of technical, regulatory, and social factors. In the context of Saudi Arabia—a country pursuing rapid digitalization,

large-scale infrastructure projects, and ambitious mobility initiatives—there is a clear need for a transparent, multi-dimensional readiness framework tailored to the national context and capable of integrating expert judgments, policy documents, and empirical data.

This research responds to that need by proposing and applying a reproducible, mixed-methods readiness assessment for AV adoption in Saudi Arabia. The framework integrates five complementary dimensions and converts coded qualitative and quantitative evidence into dimension scores and an overall readiness index using explicit, confidence-adjusted scoring rules. We apply the method to an evidence base composed of expert interviews, surveys, policy and technical documents, and published case studies, and we document a traceability matrix that links each readiness rating back to its evidence sources.

This research aims to develop and apply a transparent, reproducible readiness assessment model for AV adoption in Saudi Arabia by integrating qualitative and quantitative evidence across infrastructure, technology, data connectivity, regulatory policy, and societal market dimensions. The study further seeks to generate a cross-dimensional profile of readiness that not only identifies the overall level of preparedness, but also the specific strengths, gaps, and interdependencies shaping deployment feasibility. In doing so, the paper contributes a policy-relevant framework that can support evidence-based decision-making, regulatory planning, and future readiness monitoring in Saudi Arabia and other comparable national contexts.

The research questions include the following:

- RQ1: What is the current state of readiness for implementing AVs in Saudi Arabia?
- RQ2: How can a technology readiness assessment framework be developed to evaluate and improve Saudi Arabia's readiness for adopting AV?
- RQ3: How can the reliability and robustness of the developed framework be evaluated?

The remainder of the paper is structured as follows: Section 2 reviews related work on AV readiness and identifies the methodological gaps addressed by this study. Section 3 details the research methodology, including data collection, data analysis, framework development, and system validation. Section 4 reports the assessment results and cross-country comparability analysis. Section 5 discusses policy implications and provides prioritized recommendations. Section 6 concludes with limitations and future research directions.

2. Literature Review

Several frameworks exist for evaluating the maturity of emerging technologies, including the Technology Readiness Level (TRL), Technology Readiness Index (TRI), Technology Readiness Assessment (TRA), Autonomous Vehicles Readiness Index (AVRI), and other tailored assessment frameworks specifically designed for AVs.

The TRL serves as a standard method to evaluate how ready a particular technology is for its intended use in end products and systems. NASA created this scale during the 1970s and established it as a formal standard in 1989 to evaluate spacecraft design readiness. The original definition used seven levels, but NASA adopted the nine-level scale during the 1990s, which has become the standard for assessment [7]. Organizations, together with companies and governments, use this framework to evaluate technological development at a basic level and reach TRL 9 for the most advanced stage. The evaluation process for technology maturity stages: The achievement of TRL 5 requirements does not automatically advance the technology to the next level [8]. The TRL framework for AVs focuses on establishing standards and best practices because the regulatory framework for AVs continues to evolve [9]. The study [10] applied the TRL framework to evaluate AV readiness and demonstrated that theoretical progress exceeds current implementation

capabilities. The study found that the worldwide objective for AV integration exceeds the current state of readiness for actual deployment.

The TRI functions as a diagnostic instrument that evaluates people's readiness to use new technological systems [11]. The assessment tool evaluates people's behavioral responses through four essential categories, including their positive views about technology, their readiness to try new things, their discomfort levels, and their security worries about technology reliability and effects. The original TRI contained 36 assessment items, but the TRI 2.0 version reduced the number to 16 items. The TRI 2.0 version uses technology-agnostic terminology that makes it suitable for academic research and practical implementation [11]. The study [12] employed TRI to identify which elements affect user system adoption. The TRI system assesses technology readiness through four separate evaluation criteria, including optimism, inventiveness, discomfort, and insecurity.

The TRAs provide a structured and evidence-based method to assess Critical Technology Element (CTE) development readiness in acquisition programs through the use of TRLs as defined metrics [13–15]. The main objective of a TRA involves technology infusion through continuous and consistent evaluation processes throughout the technology lifecycle [15]. Risk management depends heavily on TRAs to handle the challenges that come with acquiring advanced technologies such as AVs [15]. The quality of the framework depends on its credibility, objectivity, reliability, and usefulness [13]. The framework provides an unbiased evaluation of technology development stages, which helps organizations make better choices about implementing new technologies [14].

The TRA process evaluates technologies based on particular requirements, which determine how far along a technology has progressed at its current TRL [15]. The evaluation process requires reassessment of all previously validated technologies that were developed for particular settings before their deployment in new operational environments [15]. The five-step process of the development of a high-quality TRA, which starts with creating a TRA plan and choosing team members, followed by CTE identification & assessment, then TRA report preparation, and finally using the report results [13].

The AVRI, developed by KPMG, is a reference that combines a comprehensive set of 28 factors across the AV ecosystem. These factors are combined into a score that prioritizes countries' readiness for automated vehicles. The AVRI, first established in 2018 [16] provided an initial comprehensive framework for assessing AV readiness. Subsequent editions in 2019 [17] and 2020 [18] have continued to refine and expand the scope of the index. The index assesses readiness through multiple dimensions, such as policy & legislation, technology & innovation, infrastructure, and consumer acceptance [18]. The AVRI has not been updated since 2020, which restricts its ability to accurately represent recent progress in AV readiness.

Recent studies have utilized the AVRI to evaluate industry readiness for AV deployment. The study [19] leveraged the AVRI to investigate the infrastructure readiness for AV technology. It identified critical factors such as road quality, 5G coverage, and Vehicle-to-Everything (V2X) technology. While focusing on infrastructure, the study may have overlooked other essential aspects, including consumer acceptance, regulatory frameworks, and technological advancements, which are equally crucial for successful AV implementation. Similarly, the study [20] utilized the AVRI to assess the readiness for transport automation. This study relies on comparisons with other countries based on the KPMG report, but it may not have fully captured the unique socio-economic and cultural factors influencing AV adoption. Another study [21] proposes a framework based on benchmarking with the AVRI, using it as a tool for assessing readiness levels for AVs. The study emphasizes the need for collaboration, clear roles, and effective data management to facilitate the successful adoption of AV technology.

The study by Ho et al. proposes a conceptual framework that analyzes the impacts of psychological factors, such as perceived safety and trust, as well as external factors, including public health, technology, legislative issues, and built environment perspectives. The researchers highlight that psychological elements are crucial in AV acceptance [22]. Another study by Sitinjak et al. uses a multi-faceted approach to provide a comprehensive assessment of readiness for AV integration that considers regulatory, infrastructural, and public perception aspects. The study emphasizes the need for robust safety regulations and infrastructure improvements to support AV adoption. However, its qualitative approach limits the generalizability of the findings to the broader population [23].

Lim et al. propose a framework focusing on four key areas: policy & legislation, infrastructure, technology, and public acceptance. The paper insists on developing a comprehensive framework that considers the various aspects of AV technology and its adoption [24]. On the other hand, Westerman et al. explore land use planning for AV integration, noting the lack of existing land-use planning strategies for AVs. They use a framework to assess and manage risks related to fully integrating AVs, measuring the level of intervention across four categories: fix, adaptation, transformation, and innovation. However, the study indicates a lack of existing policies or examples of risk management strategies to address the impacts of transitioning to AVs [25].

Jiang et al. employ the Analytic Hierarchy Process (AHP) to assess city readiness for CAVs. The study evaluated readiness across infrastructure, citizens, and policies & regulations. AHP provides a structured approach to decision-making by breaking down complex criteria into a hierarchical structure, allowing stakeholders to assess the relative importance of various readiness criteria through pairwise comparisons. However, the number of criteria that can be evaluated using the AHP is limited. The study emphasizes the lack of established frameworks and criteria to measure city readiness for CAVs [26].

The TRA system provides a complete assessment of technological readiness through its evaluation of operational condition [15]. The TRA system provides better results for complex real-world applications because it matches technological development to particular requirements [14]. The TRA system requires supporting data for operation, but it fails to evaluate essential design aspects, including reliability and safety [13]. The TRL system shows hardware and software development stages in operational settings, yet it does not account for human factors and environmental conditions [27]. The TRI system evaluates individual readiness for new technology adoption through assessments of optimism and discomfort levels [11]. However, the evaluation system fails to assess technical and environmental aspects that affect technology deployment. The AVRI framework evaluates national readiness for AV adoption through assessments of policy frameworks as well as technological and infrastructural capabilities [18]. However, the AVRI framework might struggle to handle detailed national regulations and social effects that occur when AVs enter public use. The TRA system stands out because it uses a detailed method to check technology readiness against particular needs in appropriate settings [15]. Our research adopted the TRA framework because it provides a complete method to evaluate new technology readiness, particularly for early adoption of AVs.

Previous studies have shown that frameworks developed to assess AV readiness often focus on specific aspects such as infrastructure, technology, or public acceptance. These frameworks fail to fully address the complexity of AV readiness. Studies using frameworks such as AVRI address technical and infrastructure readiness, but often overlook public acceptance, socio-economic factors, and regulatory frameworks, all of which are key to successful AV adoption [19–21]. This observation highlights a clear gap in existing approaches, emphasizing the need for a comprehensive framework. Building on this identified gap, the development of an AV framework seeks to provide a more nuanced and

comprehensive approach to assessing AV readiness in the Saudi Arabia. The framework aims to integrate multiple dimensions into a unified model, ensuring that all critical factors influencing AV deployment are systematically considered.

3. Methods

This research uses a mixed-methods approach, integrating quantitative and qualitative data to assess Saudi Arabia's readiness for AV technology implementation. The methodology, as shown in Figure 1, includes data collection, data analysis, framework development, and evaluation. Data was gathered from multiple sources, including expert interviews, policy and technical documents, and case studies. We also drew upon a mixed-methods public survey conducted in our previous research [28], which captured public opinion on AV technology. The present research continues the analysis using the same data. Our analysis combined thematic analysis of qualitative data with statistical and regression techniques for quantitative data. Also, Political, Economic, Socio-cultural, Technological, Environmental and Legal (PESTEL) and Strengths, Weaknesses, Opportunities, and Threats (SWOT) analyses were conducted. These insights directly informed the development of a multi-dimensional TRA framework, structured around five core dimensions: Infrastructure, Technology, Data & Connectivity, Regulatory & Policy, and Societal & Market readiness. The selection of these five dimensions was not arbitrary. It was derived using a structured TRA process informed by the preceding PESTEL and SWOT analyses, the interview findings, and the literature on AV readiness frameworks. These sources consistently highlighted the five dimensions as the most critical factors of AV deployment in the Saudi context. We therefore retained only the dimensions that were repeatedly supported across the evidence base and that were directly relevant to near-term AV implementation. This approach ensured that the framework remained both context-sensitive and aligned with the main operational and policy factors influencing readiness. In this way, the five dimensional framework reflects a synthesis of evidence rather than an expert-imposed list of factors. Also, build a smart framework tool that utilizes rule-based expert reasoning, a weighted scoring mechanism with crisp thresholds, and Large Language Model (LLM)-based document analysis. The evaluation and assessment of the system involve validation using a literature-based dataset and Monte Carlo simulation to assess system performance, reliability, and robustness.

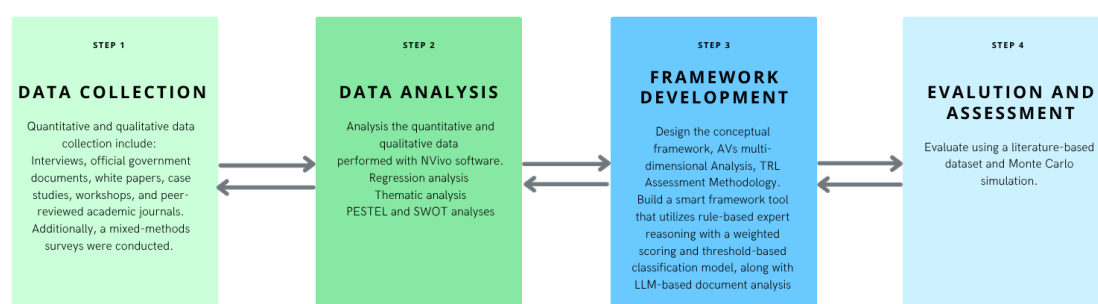


Figure 1. Flowchart of methodology.

3.1. Data Collection

Semi-structured interviews enhance the understanding of individual experiences, enabling participants to share their unique perspectives [29]. Semi-structured interviews were used with key stakeholders from different sectors related to AVs, including government sector, academic institutions, transportation and logistics, independent experts, and members of the Centre for the Fourth Industrial Revolution (C4IR). Experts were selected with 4–10 years of experience to ensure they had strong operational and strategic knowledge in the AV sector. This approach facilitated our understanding of the current state of AV technology

adoption in Saudi Arabia. This method was necessary to capture qualitative insights and expert judgments that are not available in the published literature. Government policy papers, technical reports, and white papers were thoroughly reviewed. These documents provided detailed information concerning the present and future advancements in AV technologies. The information was used to conduct a comprehensive review, revealing the multifaceted implications of AVs across various domains. Documents were selected based on their relevance to national and regional AV strategies, technological standards, and infrastructure planning. Analyzing pilot projects, case studies, and workshops was necessary to bridge the gap between theoretical models and real-world implementation. Projects and cases were selected to represent a diverse range of implementations relevant to the Saudi context and the global best practices. The review of these projects utilized publicly available reports, official evaluations, and published case study data. The research reviewed international peer-reviewed journals and academic papers to assess AVs readiness in different countries. The literature search was conducted across major academic databases like Scopus and Web of Science, using keywords related to AV readiness, technology assessment, and autonomous public transportation. This systematic review helped build the research's theoretical foundation and compare its findings with global research. We used a mixed-methods public survey from our previous research [28] to capture public perceptions of AVs in Saudi Arabia. This approach measures both quantifiable attitude trends and their underlying causes, providing a comprehensive view of AV readiness acceptance.

3.2. Data Analysis

Thematic analysis is a widely utilized approach for examining qualitative data [30]. The standard procedure includes familiarization with the data, generating initial codes, identifying themes, reviewing and refining themes, and producing a final report [31]. This method is recognized for its flexibility and its capacity to yield in-depth contextual insights into real-world phenomena, such as individual beliefs, perspectives, and experiences [32]. Thematic analysis was applied to data obtained from expert interviews. Two semi-structured interviews were conducted at different times with key stakeholders in the field. All participants ($n = 7$) were interviewed via video call. We acknowledge this sample size as a clear limitation for statistical generalization. The first interview was a group interview with six stakeholders, while the second interview was an individual interview with one stakeholder. Each interviews were conducted as one session with an average duration of one hour. The interview structure were designed with explicit consideration for the demanding schedules of domain experts. While acknowledging that the influence of groups and time constraints represent a potential limitation for the depth of knowledge, this factor was carefully considered during the subsequent analysis and interpretation of the findings. The interview data were analyzed using NVivo, a qualitative analysis software that facilitates systematic coding and organization of qualitative data. The interview data were coded by two independent coders with experience in qualitative analysis and technology assessment. Both coders used the same predefined codebook to assign statements to the five readiness dimensions and their associated factors. After the initial coding round, inter-coder agreement was checked by comparing the assigned codes across all excerpts, and disagreements were resolved through discussion until consensus was reached. When consensus could not be reached immediately, the coding decision was reviewed against the original transcript, the code definition, and the study's readiness criteria before a final joint decision was made. This procedure was used to improve consistency, reduce subjective bias, and ensure that the mapping from evidence to readiness dimensions remained transparent and reproducible. The quantitative statistical and regression techniques used to analyze

the survey data were performed in our previous research [28]. Additionally, PESTEL and SWOT analyses were conducted to provide a broader contextual understanding.

3.3. Evidence Integration and Quantification Procedure

This research adopted a structured mixed-methods procedure to convert qualitative and quantitative evidence into a transparent readiness assessment framework. The evidence integration process followed a sequence of collection, screening, coding, scoring, weighting, and aggregation. Each source of evidence was mapped to one of five readiness dimensions using a predefined codebook. The coded items were then converted into ordinal scores, adjusted for evidence confidence, and aggregated into dimension-level and overall readiness scores. The final dimension and overall scores were translated into TRL value using predefined threshold rules.

Evidence integration workflow:

1. Collect evidence from interviews, surveys, PESTEL, SWOT, the literature, and expert judgments.
2. Screen and clean the material to remove irrelevant, duplicate, or unsupported statements.
3. Code each item into one of the five readiness dimensions using a predefined codebook.
4. Convert coded items into scores using a fixed ordinal scale, such as 0 = absent, 1 = partial, and 2 = strong evidence.
5. Assign confidence levels to reflect how strong, consistent, and traceable each piece of evidence is.
6. Aggregate factor scores into dimension scores using weighted averages.
7. Combine dimension scores into an overall readiness index.
8. Map the final index to the corresponding TRL or readiness band using predefined thresholds.
9. Validate the result by checking expert agreement, evidence traceability, and consistency across sources.

The evidence base consisted of semi-structured expert interviews, survey findings, PESTEL and SWOT analyses, policy and technical documents, the published literature, and domain expert judgments. These sources were first screened, coded, and mapped to five predefined readiness dimensions: Infrastructure, Technology, Data & Connectivity, Regulatory & Policy, and Societal & Market readiness.

Qualitative data from interviews and documents were analyzed using a structured content-analysis procedure to identify readiness-related statements, recurring constructs, and context-specific indicators. Each extracted statement was then assigned to one or more readiness dimensions using a predefined codebook. To ensure consistency, only statements that directly reflected operational capability, policy maturity, implementation feasibility, user acceptance, or data governance were retained as scoring evidence. Ambiguous statements were excluded from the quantitative layer and retained only as supporting context.

To enable quantification, each factor within a readiness dimension was converted into an ordinal maturity score using a three-level scale: 0 = absent or not evidenced, 1 = partially evidenced or emerging, and 2 = clearly evidenced or mature. Where quantitative survey results were available, their outcomes were normalized to the same scale before integration. When multiple evidence sources supported the same factor, the factor score was computed as the mean of the source-specific scores, adjusted by an evidence-confidence coefficient reflecting source agreement, data completeness, and expert verification. This confidence-weighted averaging served as the explicit mechanism for resolving contradictions between sources: rather than selecting one source over another, disagreeing sources were combined in proportion to their confidence weighting, so no single conflicting source could dominate the final score.

Dimension scores were calculated as weighted aggregates of the factor scores within each dimension. Equal weights were applied to factors unless a factor was judged to have disproportionate importance based on expert consensus and repeated support across the literature. The general formula used is

$$D_j = \frac{\sum_{i=1}^n w_i x_{ij}}{\sum_{i=1}^n w_i} \quad (1)$$

where D_j is the score of dimension j , x_{ij} is the score of factor i within that dimension, and w_i is the assigned weight of factor i . The overall readiness score was then derived from the weighted average of the five dimension scores:

$$R = \frac{\sum_{j=1}^5 W_j D_j}{\sum_{j=1}^5 W_j} \quad (2)$$

where R is the overall readiness score and W_j is the weight of dimension j . In the base model, equal dimension weights were used to preserve methodological neutrality and to avoid overemphasizing any single domain.

To translate readiness scores into TRL values, crisp thresholds were defined a priori. Each dimension score and the overall readiness index were classified into maturity bands aligned with the TRL logic used in technology assessment. Higher scores indicated stronger evidence of implementation preparedness, while lower scores reflected conceptual, experimental, or fragmented readiness. Where evidence was inconsistent or incomplete, a confidence adjustment was applied to prevent overstating readiness. This adjustment reduced the final score proportionally to the degree of uncertainty and evidence disagreement.

A traceability matrix, shown in Table 1, was maintained throughout the analysis to document the source of each coded item, its assigned factor, the relevant readiness dimension, the score applied, and the rationale for the decision. This audit trail allowed for direct tracing from raw evidence to final readiness classification and supported transparency, reproducibility, and methodological accountability. The final scoring model was reviewed by domain experts to ensure that the interpretation of evidence remained grounded in the Saudi AV context and aligned with established readiness assessment principles. The complete dataset comprises 2 semi-structured interviews, 1 survey with 208 respondents from Saudi Arabia [28], 1 PESTEL analysis, 1 SWOT analysis, 53 peer-reviewed literature sources, and 20 government policy papers, technical reports, and white papers. All data were collected between October 2024 and December 2025. All secondary sources are cited in the reference list; primary interview and survey data are proprietary to the authors' research and not publicly available due to participant confidentiality.

The following are the scoring rules:

- Score 0 = no evidence, weak evidence, or only conceptual mention.
- Score 1 = partial, emerging, or inconsistent evidence.
- Score 2 = clear, repeated, and verifiable evidence.
- Confidence adjustment = applied when sources conflict, data are incomplete, or expert consensus is weak.
- Final TRL = determined from the weighted readiness index using predefined thresholds.

Table 1. Evidence to readiness scoring matrix.

Source Type	Evidence Item	Coded Factor	Readiness Dimension	Raw Score (0–2)	Weight	Confidence Level	Confidence Coefficient	Adjusted Score	Justification
Interview	Expert statement on V2X pilots	V2X implementation	Infrastructure readiness	2	1.0	High	1.0	2.0	Direct operational evidence
Survey	Public trust in AVs	User acceptance	Societal Market readiness	1	1.0	Medium	0.8	0.8	Positive but not mature
PESTEL	Regulatory sandbox reference	Policy flexibility	Regulatory Policy readiness	1	1.0	High	1.0	1.0	Formal policy support exists
SWOT	Weak data-sharing capacity	Data governance gap	Data Connectivity readiness	0	1.0	High	1.0	0.0	Clear weakness in evidence base
Literature	Controlled testing reported	Prototype validation	Technology readiness	2	1.0	High	1.0	2.0	Repeated in peer-reviewed studies

LLM is used as a supporting coding assistant with human-in-the-loop verification, while the actual readiness score is computed by explicit equations and fixed rules. For evidence confidence, we use

$$x'_{ij} = x_{ij} \times c_{ij} \quad (3)$$

where x'_{ij} is the adjusted factor score and c_{ij} is the confidence coefficient, for example, 1.0 for high confidence, 0.8 for medium confidence, and 0.6 for low confidence.

While TRLs have been extremely successful in areas such as aerospace and the defense sector, their straightforward application in the domain of AVs is constrained by multiple limiting factors. One important limiting factor is system complexity. AVs are not merely a technology, but a combination of various subsystems. Therefore, the readiness assessment must include the maturity of each subsystem and the reliability of its integration as well [9]. To reflect this multi-system nature of AV readiness while retaining continuity with the classical TRL vocabulary, this study adopts a modified TRL scale [8] comprising eight readiness bands, applied consistently across all five readiness dimensions (shown in Table 2).

Table 2. Readiness score to TRL mapping using the modified TRL scale (adopted from [8]).

Readiness Score Range	TRL Interpretation	Description
0.000–0.124	TRL 1–2	Basic principles observed
0.125–0.249	TRL 2–3	Technology concept formulated
0.250–0.374	TRL 3–4	Proof of concept
0.375–0.499	TRL 4–5	Component validation in laboratory
0.500–0.624	TRL 5–6	Technology validated in relevant environment
0.625–0.749	TRL 6–7	System demonstrated in relevant environment
0.750–0.874	TRL 7–8	System demonstrated in operational environment
0.875–1.000	TRL 8–9	System proven through deployment

3.4. Multi-Dimensional TRA Analysis

The TRA dimensions of the framework are: Infrastructure readiness, which includes availability of infrastructure, 5G and spectrum policy, V2X activity, Saudi Highway Code (SHC) 801 codification, and Electric Vehicle Infrastructure Company (EVIQ) charging expansion; Technology readiness, involving real-world controlled testing beyond prototypes, local industrial roadmaps, and addressing gaps in safety cases; Data & Connectivity readiness, which focuses on connectivity infrastructure, developing data governance frameworks, and establishing cross-border data flow and liability backbone;

Regulatory & Policy readiness, which encompasses regulatory sandboxes, policy visions for AVs, ongoing AV pilots, and frameworks for licensing, liability, and insurance at the national level; and Societal & Market readiness, which assesses early-stage public readiness, positive trends in acceptance, trust-building pilots, outreach and surveys, and identifies clear near-term use cases.

3.5. Cross-Country Comparability and Framework Novelty

The Saudi Arabian evidence base is partly comparable with other countries, but not directly equivalent in all respects. It is comparable at the level of broad readiness dimensions, because the same framework evaluates infrastructure, technology, data connectivity, regulation, and societal market conditions across cases. However, it is not fully comparable at the level of raw indicators, because countries differ in data availability, policy maturity, urban form, pilot project scale, traffic conditions, digital infrastructure, and public acceptance patterns.

For that reason, comparisons should not be made using final readiness scores alone. Instead, the analysis should compare the underlying conditions that generated those scores, as captured by the specific factors scored within each dimension of the framework. For example, in the regulatory dimension, the relevant evidence includes the maturity of regulatory sandbox mechanisms, the extent to which national AV policy vision has been codified into formally enacted legal frameworks, the scale and strategic scope of ongoing pilot projects, and the development stage of a comprehensive national framework addressing licensing, liability, and insurance. In the infrastructural dimension, the dataset should reflect the deployment stage of V2X communication technology, the adequacy of physical infrastructure for AV operation, the development of 5G connectivity and dedicated spectrum allocation, the formal codification of relevant national standards (such as SHC 801), and the progress of electric vehicle charging infrastructure expansion.

This means that two countries may produce similar readiness scores while still being different in substance. One country may score highly because of strong policy alignment but limited pilot deployment, while another may score moderately because of advanced testing infrastructure but weak governance or lower public acceptance. The framework treats readiness as a multi-dimensional profile rather than a single ordinal ranking, which makes the interpretation more policy-relevant than a simple score comparison.

Methodologically, the framework differs from many existing readiness assessment approaches in three ways. First, it integrates qualitative and quantitative evidence within a single transparent scoring pipeline rather than relying only on expert judgment or only on numerical indicators. Second, it applies confidence-adjusted scoring, so weak or partially supported evidence does not carry the same weight as direct and verifiable evidence. Third, it preserves traceability through a source-to-score audit trail, allowing readers to see how each piece of evidence contributed to the final readiness result.

This approach is especially important when assessing Saudi Arabia because the national context includes rapid digital transformation, evolving regulatory structures, and uneven maturity across technical and social domains. A framework that simply compares final scores would miss these differences.

In comparison with conventional technology readiness or maturity models, the present framework is broader and more context-sensitive. Traditional models often emphasize technological development stages, whereas this study also incorporates regulatory readiness, data governance, and societal acceptance as equal determinants of deployment feasibility. As a result, the framework is better suited to national policy planning, cross-country benchmarking, and regulatory prioritization.

3.6. System Architecture

Figure 2 presents a framework for a hybrid expert system that integrates rule-based reasoning, a weighted scoring model with threshold-based classification, and LLM support for document analysis and rule generation. In the mandatory workflow, Users complete multi-dimensional questionnaires. The system then calculates a TRL score using existing rules and the weighted scoring mechanism. In the optional workflow, users can upload a SWOT or PESTEL file, or both, which the LLM analyzes to propose new rules for the knowledge base. The LLM extracts candidate rule specifications, such as factor weights, thresholds, or new indicators, and presents them in a structured list to the human reviewer, who accepts, modifies, or rejects each proposal. Only accepted rules are committed to the knowledge base, after which the TRL is automatically recalculated. The TRL is recalculated after each approved update. A human-in-the-loop validation step determines which LLM proposed rules are accepted into the production knowledge base.

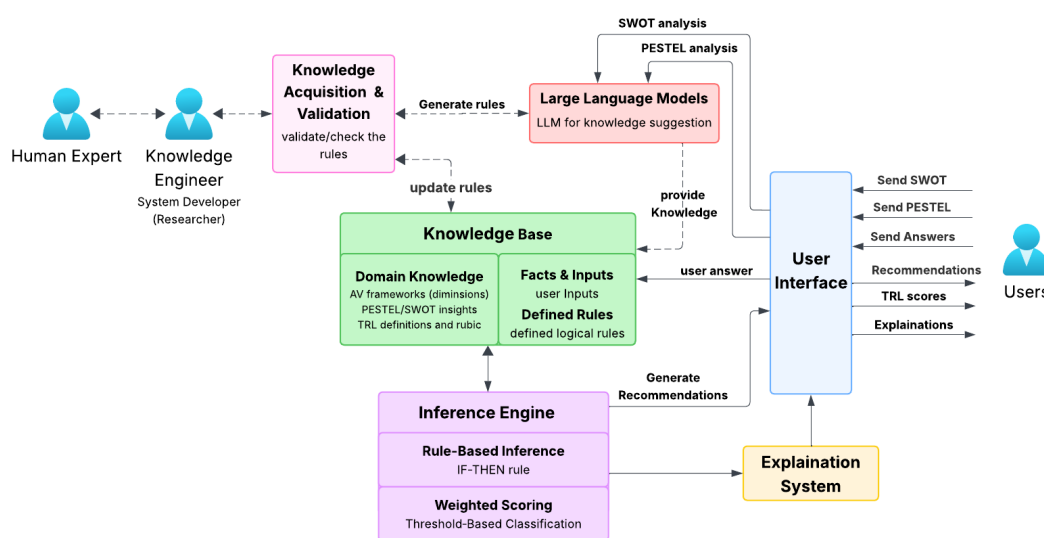


Figure 2. Framework architecture.

3.7. System Validation Approach

The system was validated using literature-based datasets and Monte Carlo simulations. The literature-based validation dataset utilizes academic and industry sources, including KPMG AVRI [16,17], the MDPI AV implementation in developing countries [33], and the Tenet AV Readiness Ranking [34]. The research also used the analysis of the current state in Saudi Arabia in the dataset. These sources evaluate 20 regions, each assessed across multiple dimensions of AV readiness using established frameworks. The compilation of validation cases necessitated a systematic alignment of source assessments with the present research five dimensional framework. Each source utilized distinct conceptual frameworks: KPMG structured its assessments around policy, technology, infrastructure, and consumer acceptance [16,17], MDPI focused on physical and digital infrastructure and societal perceptions of AV [33], and Tenet's evaluations concentrated on infrastructure provision metrics [34]. The alignment process involved systematic mapping of source assessment dimensions into the research framework's five dimensions.

The Monte Carlo simulation was designed to stress-test the system's stability and robustness by evaluating its performance across thousands of randomly generated assessment scenarios [35,36]. A multistep validation process using Monte Carlo simulations confirms system accuracy by testing performance across a variety of conditions with different parameter combinations [37,38]. Monte Carlo simulation was chosen as the validation method because it is the only practical way to assess complex, non-linear systems when

analytical solutions are not possible. It also generates sampling distributions that show system behavior across all possible input conditions [39]. The process employs stratified sampling to ensure balanced distribution across all five readiness dimensions and three maturity levels. The simulation was configured with 500 default scenarios and response consistency across the defined Nascent, Developing, and Mature input states. A fixed random seed (42) was used to ensure reproducibility. For each simulation run, random maturity levels are generated for each input factor across all five dimensions. The overall readiness score is then recomputed for each scenario using the same weighted aggregation procedure described in Section 3.3. This approach generates a distribution of possible readiness outcomes, enabling us to assess the stability and sensitivity of the framework to variations in input maturity levels.

3.8. Validation Metrics and Success Criteria

The validation framework employs multiple complementary metrics to assess system performance across different dimensions of accuracy and reliability. The literature agreement rate measures how well system predictions align with established expert benchmarks from the published literature. For this assessment, we define an agreement rate greater than 80% as the threshold for high system accuracy. The ± 1 TRL band represents a reasonable tolerance reflecting the inherent uncertainty in readiness assessment while still ensuring magnitudes remain reliable for decision-making.

$$\text{Agreement Rate} = \frac{\text{Num of Predictions Within } \pm 1 \text{ TRL of Literature}}{\text{Total Predictions}} \quad (4)$$

The Mean Absolute Error (MAE) metric provides complementary information to the agreement rate by quantifying the average magnitude of prediction errors across all validation cases. This metric proves particularly useful for identifying systematic biases in system performance. We establish a MAE of ≤ 1.0 as the benchmark for good performance in this system. For MAE calculation, each TRL band is converted to its numeric midpoint. Both the system's predicted band and the expert-assigned ground-truth band are converted accordingly.

$$\text{MAE} = \frac{1}{\text{Num of Predictions}} \times \sum |\text{Predicted TRL} - \text{Actual Ground Truth}| \quad (5)$$

The Root Mean Square Error (RMSE) metric emphasizes larger errors while remaining less sensitive to outliers compared to maximum error metrics. This metric is particularly useful for identifying problematic variability in prediction accuracy, with some subset of cases generating substantially larger errors than typical predictions:

$$\text{RMSE} = \sqrt{\frac{1}{\text{Num of Predictions}} \times \sum (\text{Predicted TRL} - \text{Actual Ground Truth})^2} \quad (6)$$

The Stability Rate measures internal consistency and boundary behavior when processing randomly generated input combinations with systematic perturbations, validating operational reliability. This metric validates that the system behaves reliably at decision boundaries and does not produce erratic responses to small input variations. We define a Stability Rate of $\geq 90\%$ as indicative of production readiness.

$$\text{Stability Rate} = \frac{\text{Number of Stable Scenarios}}{\text{Total Scenarios}} \times 100 \quad (7)$$

Model validation was conducted using benchmark cases from the literature and published AV readiness references, together with controlled perturbation testing. The proposed

framework was applied to selected cases representing different readiness conditions, and the resulting TRL outputs were compared with expected or expert-derived benchmark TRL levels reported in the source studies. In addition, a Monte Carlo simulation was used to test the robustness of the weighted scoring model under random variation in factor weights and evidence-confidence values. The validation results indicated that the model produced stable outputs across repeated runs and remained sensitive to meaningful changes in the input structure without generating inconsistent TRL classifications. This supports the reliability and practical applicability of the proposed readiness assessment framework.

4. Results

4.1. Overview of the Framework Application

The assessment methodology uses multiple evaluation methods to assess AV readiness through five essential dimensions, including Infrastructure readiness, Technology readiness, Data & Connectivity readiness, Regulatory & Policy readiness, and Societal & Market readiness. Each dimension is assessed independently using a standardized questionnaire tool designed to capture the maturity level of specific factors in that dimension.

The assessment tool uses a three-point Likert scale to measure readiness levels for each factor within its dimensions. The scale starts at one (Nascent), showing the beginning of development with no significant advancement, then proceeds to two (Developing), indicating moderate progress with partial deployment and pilot programs getting underway, and ends at three (Mature), indicating fully operational systems that have achieved widespread deployment. The selected ordinal scale follows industry-standard TRL frameworks while remaining easy for domain experts and policy stakeholders to understand. TRL provides a systematic metric for assessing technology maturity on a scale from TRL 1 (basic principles) to TRL 9 (proven system). For AV applications, we adopt and modify the scale to align with AV-specific terminology and development stages.

For each dimension, specialized questionnaires target the factors most relevant to that domain's contribution to the overall AV readiness. For instance, the Infrastructure readiness dimension evaluates five factors: V2X communication maturity, physical infrastructure adequacy, 5G connectivity advancement, standards codification (SHC 801), and EV charging infrastructure progress.

The questions were validated using the content validity method with two subject matter experts rating the relevance of each item on a four-point Likert scale, in which "1 = Not Relevant; 2 = Somewhat Relevant; 3 = Quite Relevant; and 4 = Highly Relevant. The overall average was 98% (S-CVI = 0.98), which is an acceptance of validity based on [40,41].

The overall TRL assessment is calculated by combining individual dimension assessments, where each one receives its weight from confidence scores. The scoring system uses a weighted average approach, which assigns values to each factor within its dimension. The system applies equal weight to all factors. The completeness of responses within each dimension, together with the consistency of maturity levels across factors, determines the accuracy of confidence scores. The system performs calculations for each of the five dimensions by multiplying the numeric TRL values by their confidence scores, then dividing the results by the overall confidence weight. The system generates a TRL score based on the weights, which determines the correct TRL band.

The assessment process faces major uncertainty that arises, as stakeholders may provide inaccurate ratings for particular factors, which prevents them from providing reliable ratings for certain factors. The system uses a hierarchical fallback system to manage situations where assessments remain inaccurate. The knowledge engine requires more data than is available to perform rule-based reasoning. The system shown in Figure 3

uses a pattern-matching and similarity-based method to estimate missing values, drawing from answered factors and similar cases found in the validation database. The assessment process continues without interruption through this method, which shows the confidence level for each assessment dimension.

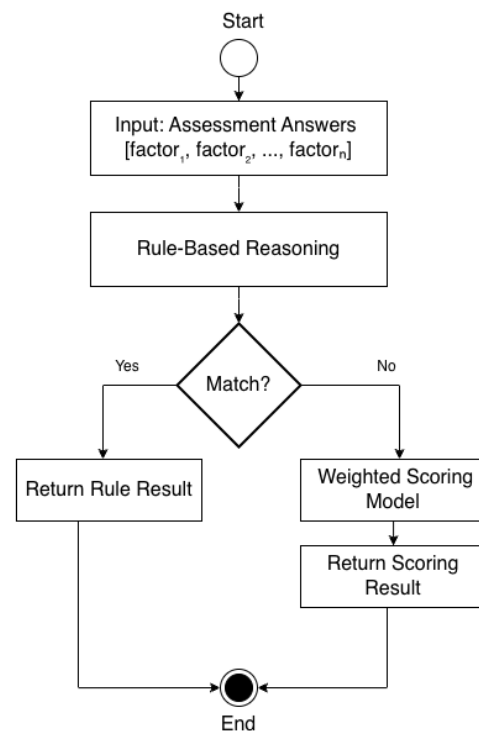


Figure 3. Hybrid expert system decision workflow.

4.2. Data Presentation

The dataset exhibits significant geographic diversity, comprising 34 case studies from 20 nations, shown in Figure 4. This includes seven European countries, six from the Asia-Pacific region, four from the Americas, two from the Middle East, and one from Africa. It is important to note that five case studies are drawn from Saudi Arabia, three case studies from Singapore, the United States, and the United Arab Emirates, and two case studies from the Netherlands, Sweden, Germany, and Japan, providing a more in-depth perspective from these specific contexts.

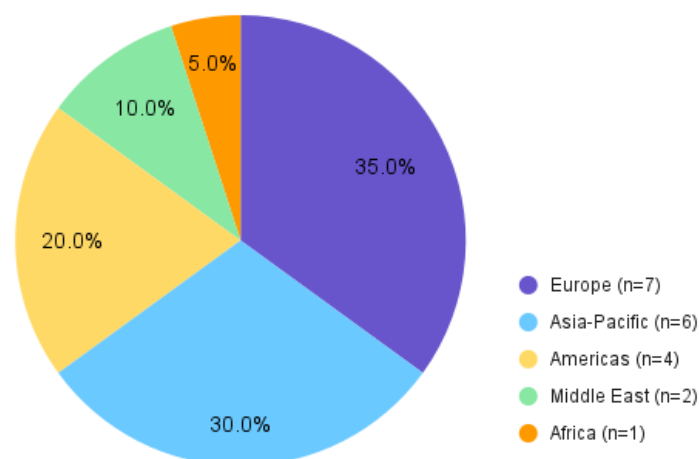


Figure 4. Geographic diversity of the dataset [16,17,33,34].

The dataset is well-distributed across the full spectrum of TRLs. Five cases represent TRL 8-9, primarily from developed nations with mature AV ecosystems and regulatory frameworks. Eight cases represent TRL 7-8, reflecting a system with demonstrated capability in relevant environments. Eleven cases represent TRL 6-7, indicating substantial development underway with positive validation results. Three cases represent TRL 5-6, reflecting mid-stage development and partial validation. Seven cases represent TRL 4-5 or lower, representing nascent-stage development. This distribution provides adequate validation data across all TRL ranges.

The number of cases analyzed for each readiness dimension as following: Infrastructure readiness was the most frequently assessed dimension, comprising 25.8% of the total cases ($n = 8$), followed closely by Technology and Data & Connectivity readiness, both at 20% ($n = 7$). In contrast, Regulatory & Policy and Societal & Market readiness were assessed less frequently, each representing 17.1% of cases ($n = 6$).

4.3. Literature-Based Results

The validation results of the literature-based validation demonstrate that, as shown in Figure 5, a strong linear correlation is observed between the system's predictions and the expected TRL values from the literature. The model achieves a low MAE of 0.44 TRL levels. This indicates an average error of less than half a TRL level. Analysis by dimension shows the system achieves high agreement with literature benchmarks >80% for Infrastructure, Technology, Regulatory, and Data & Connectivity dimensions. The Societal & Market dimensions show more variable agreement.

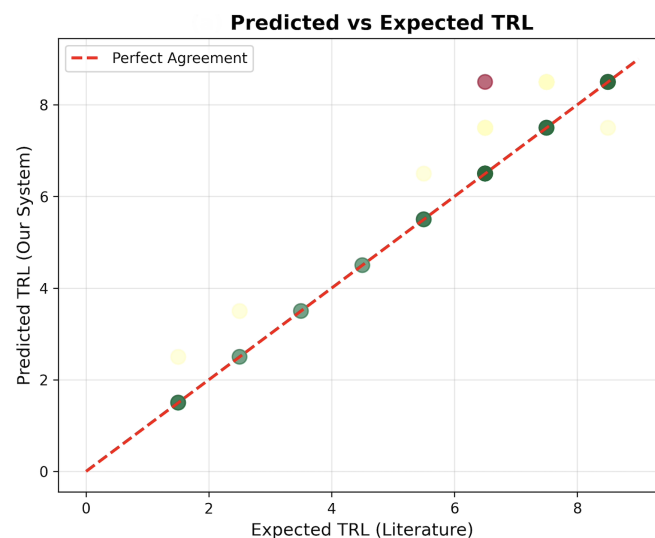


Figure 5. Literature-based validation analysis. Note: Point color indicates prediction error magnitude on a green-to-red scale: green/dark green = low error; yellow = moderate error; red/dark red = largest deviations. Red dashed line denotes perfect agreement.

4.4. Case Study Finding: AVs in Saudi Arabia

The results, shown in Figure 6, demonstrate the system's ability to generate a nuanced readiness profile for Saudi Arabia and pinpoint specific strengths and weaknesses. The system's overall predicted TRL for Saudi Arabia's AV readiness was 7-8. A comparison plot shows a close alignment between the predicted and expected TRL derived from the literature. Saudi Arabia's TRL levels are most advanced in the Technology, Infrastructure, and Data & Connectivity dimensions. To make the cross-country comparison explicit, Saudi Arabia was benchmarked against the strongest country observed in each readiness dimension rather than against a single overall leader. Singapore served as the benchmark

for Infrastructure, Data & Connectivity, and Regulatory dimensions, the United States for Technology dimension, and Sweden for the Societal & Market dimension. This dimension-by-dimension comparison was used to identify where Saudi Arabia is already approaching leading international performance and where the largest gaps remain. The comparison also shows that Saudi Arabia is relatively advanced in Technology, Infrastructure, and Data & Connectivity dimensions, while regulatory maturity and societal acceptance still lag behind the top-performing countries. This approach provides a more meaningful comparison than a single aggregate ranking, as the AV readiness develops unevenly across dimensions. This comparison validates the system's ability to generate strategic insights through cross-national comparison.

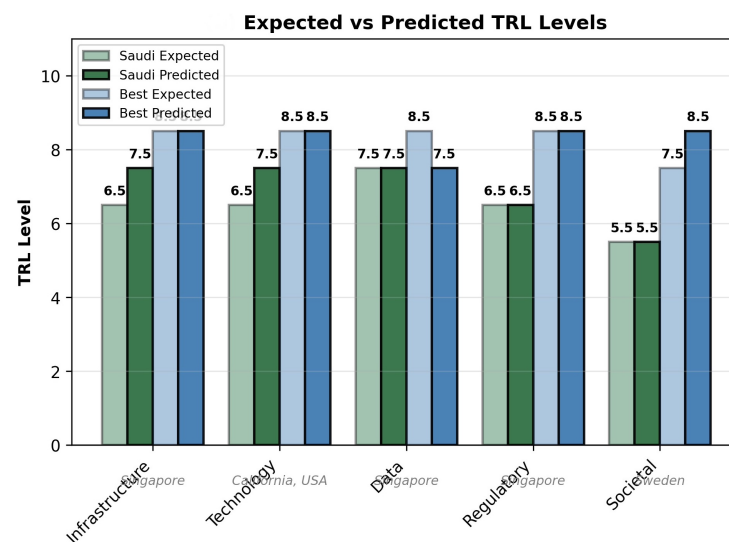


Figure 6. Comparative analysis with leader countries. Note: Light green/dark green= Saudi Arabia (expected/predicted); light blue/dark blue = best-performing country from the literature (expected/predicted). A color legend is also provided in the top-left corner of the figure.

4.5. Model Validation and Reliability

The robustness of the final aggregation and classification procedure was evaluated using a Monte Carlo simulation. Two validation approaches were used: extended simulation tested performance across 500 random scenarios, and deterministic validation tested stability by introducing small input changes. These confirmed that the system performed well and remained stable. To evaluate the overall system behavior, 500 random scenarios were generated across all five readiness dimensions. This exploration characterized the system's prediction patterns, confidence levels, and reasoning methods across diverse input configurations. The system achieved 88.3% average confidence across all scenarios. To assess production readiness, 5000 scenarios were subjected to systematic perturbation testing. Each factor was incrementally adjusted to measure output sensitivity and boundary stability. The system achieved a 92.1% stability rate.

To validate the LLM-based document analysis and rule generation module, a controlled test was conducted using Google's Gemini LLM. The test was designed to simulate a typical user interaction where external documents are uploaded for analysis. For this purpose, synthetic test documents ($n = 6$) were generated based on published literature sources representing distinct geographic and thematic focuses. The LLM processed these documents to propose readiness assessments, and the resulting TRL predictions were then compared against expert-derived benchmarks to evaluate the accuracy and reliability of this automated workflow. The validation results demonstrate that a clear positive correlation exists between the system-predicted TRL and the expected TRL values from the literature.

5. Discussion

This research aimed to develop and validate a comprehensive framework for assessing AV readiness in Saudi Arabia. The hybrid expert system brought together key factors into a clear assessment model, which was validated using a benchmark dataset and simulation inputs. The research not only met its original goals, but also provides a practical tool for policymakers, industry leaders, and researchers by adding valuable insight to the technology readiness assessment landscape. The validation results demonstrate the strong performance of the TRL assessment system. The high predictive accuracy and consistency across countries validate the system's robustness across different geographic contexts. The positively skewed error distribution and low MAE of 0.44 confirm the model's reliability. The high agreement (>80%) on key technical and infrastructural dimensions validates the system's accuracy on those core conceptual dimensions. The more variable agreement on the Societal & Market dimension can be attributed to it being assessed on fewer cases, suggesting a need for more data or highlighting the challenge of standardizing such assessments.

The results demonstrate the system's ability to generate a nuanced readiness profile and pinpoint specific strengths and weaknesses for a country. The analysis reveals a multi-faceted readiness profile for Saudi Arabia, showing a pattern of strengths and weaknesses. Saudi Arabia's AV ecosystem demonstrates a promising but uneven readiness profile, with an overall TRL of 7-8. This indicates a successful demonstration in controlled environments. The nation excels in Data & Connectivity readiness (TRL 7-8), driven by world-class 5G infrastructure. Significant progress is also evident in Infrastructure (TRL 7-8) with robust physical roads and EV networks, however. Similarly, Technology readiness (TRL 7-8) is maturing through active robotaxi pilots and major manufacturing investments, yet requires safety validation and expanded operational domains. Regulatory & Policy readiness (TRL 6-7) is maturing through sandboxes and pilot licenses, but lacks comprehensive liability and data-sharing standards. The most significant bottleneck is Societal & Market readiness (TRL 5-6). This research shows that Saudi Arabia has advanced Technology, Infrastructure, and Data & Connectivity readiness, but its Regulations and Societal readiness are still catching up. However, developing strong policies and earning public support are taking longer. This research aligns with previous research in identifying societal acceptance as a key constraint on the adoption of AVs [10,21].

The system's operational stability was evaluated using a Monte Carlo simulation, which generated hundreds of random input combinations to stress-test the internal consistency of the scoring logic. The stability of results provides statistical confidence in the system's reliability. The 92.1% stability rate shows that the system predictions are robust and stable, even with imperfect or conflicting data. Figure 7 shows the distribution of boundary stability rates across key dimensions, with Regulatory & Policy and Societal & Market readiness being the most stable and Technology readiness being the least stable.

Regulatory & Policy readiness and Societal & Market readiness exhibited the most robust internal consistency during simulation. Its highest stability rate of 100% indicates that the system's own rules and logic produce highly self-consistent outputs for this dimension across a wide array of random inputs. Infrastructure readiness demonstrates strong internal stability, with a 95.9% stability rate; the Data & Connectivity readiness displayed similar robustness, showing an 83.1% stability rate. This shows that the system's method for calculating infrastructure scores is highly stable and repeatable. Technology readiness reveals the greatest internal sensitivity during stress-testing. The stability rate of 81.7% suggests that the system's logic for this dimension is more sensitive to specific input combinations. This indicates that the rules for assessing core technology maturity are the

most likely to produce variable outputs from similar inputs. This does not mean that the answers are wrong, but that the system's internal state is less consistent for this dimension.

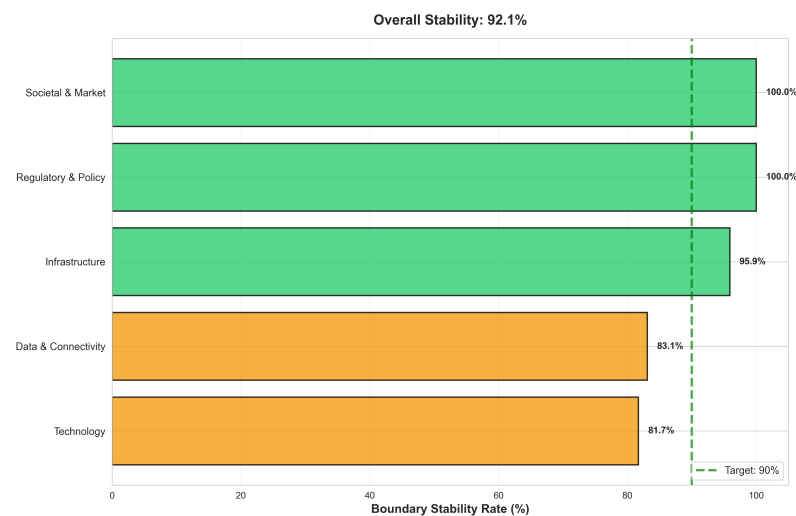


Figure 7. Boundary stability by dimension. Note: Boundary stability is the proportion of tested scenarios for which perturbing any single input factor by ± 1 maturity level changes the output readiness band by no more than two classification levels. A tested scenario is one complete input profile drawn via stratified sampling. For each scenario, each factor is perturbed individually by ± 1 level while the others remain fixed. A scenario is stable if the largest TRL change across all perturbations is $\leq$ two levels.

The exploratory validation of the LLM-based document analysis module confirmed its potential as a viable pathway for future system extensibility. The validation results demonstrate a strong overall performance for the LLM-supported workflow. The fact that both analytical frameworks (PESTLE and SWOT) performed within an acceptable band confirms the versatility and reliability of the underlying LLM-assisted coding approach for different analytical lenses. The LLM module is used as a supportive text-analysis component rather than as a standalone source of truth or as the primary computational engine of the framework. The positive outcomes from this preliminary investigation suggest that automated interpretation of unstructured texts is a feasible approach for enhancing the system's knowledge acquisition capabilities, positioning it as a promising feature for future development. However, the LLM module outputs may be affected by bias, ambiguity, and inconsistent interpretation across heterogeneous documents, which is why human validation remains essential before rule adoption.

The system's assessments provide a validated benchmark against global leaders and a strategic profile for Saudi Arabia. As shown in Table 3, its evaluations align closely with established benchmarks for leading nations: it correctly identifies Singapore's mature infrastructure and regulatory readiness (TRL 8-9) and the United States' advanced Technology readiness (TRL 8-9). A notable divergence, shown in Table 4, occurs for Singapore's Data & Connectivity, where the system assigns TRL 7-8 against a reported benchmark in the literature of TRL 8-9. This lower rating may stem from the system's more conservative evaluation of operational integration, despite major investments in 5G and smart infrastructure, which are measurable indicators that the literature may weigh differently. For Sweden's Societal & Market readiness, the system assigns a higher TRL 8-9 than the literature (TRL 7-8), likely reflecting strong, measurable indicators of public acceptance and market maturity that the academic literature may not yet fully capture.

Table 3. TRL assessment: Saudi Arabia vs global benchmarks.

Dimension	Country	Expected TRL (Literature)	Predicted TRL (System)	Error	Confidence Rate
Infrastructure Readiness	Saudi Arabia	TRL 6–7	TRL 7–8	1	92.0%
	Singapore	TRL 8–9	TRL 8–9	0	95.0%
Technology Readiness	Saudi Arabia	TRL 6–7	TRL 7–8	1	90.7%
	United States	TRL 8–9	TRL 8–9	0	95.0%
Data & Connectivity Readiness	Saudi Arabia	TRL 7–8	TRL 7–8	0	90.7%
	Singapore	TRL 8–9	TRL 7–8	1	95.0%
Regulatory & Policy Readiness	Saudi Arabia	TRL 6–7	TRL 6–7	0	90.0%
	Singapore	TRL 8–9	TRL 8–9	0	95.0%
Societal & Market Readiness	Saudi Arabia	TRL 5–6	TRL 5–6	0	84.0%
	Sweden	TRL 7–8	TRL 8–9	1	95.0%

Note: Confidence rate reflects the system’s internal certainty in its prediction. Confidence rate is calculated as 0.95 for exact expert system rule matches, or $\min(0.95, 0.6 \times \text{Completeness} + 0.4 \times \text{Maturity})$ otherwise. Completeness is the proportion of answered questions in that dimension. Maturity is the weighted average score of answered factors, divided by 3 and capped at 1. The weights (0.6, 0.4) were set heuristically to prioritize evidence completeness over score consistency and were not empirically calibrated.

For Saudi Arabia, the analysis reveals a targeted development path. The system assigns a higher TRL 7–8 to Infrastructure and Technology than literature benchmarks (TRL 6–7), likely recognizing major investments in 5G, pilot projects, and industry partnerships. This suggests Saudi Arabia is catching up with top nations in these areas. However, for Regulatory and Societal readiness, the system’s scores (TRL 6–7 and 5–6) are the same as those in the literature, showing that these areas are still the main challenges [10,20–22]. Consequently, advancing toward full commercialization is contingent upon improving these social and institutional factors.

Table 4. Notable divergences in TRL assessment.

Dimension	Country	Reason for Divergence
Infrastructure	Saudi Arabia	The system assigns a higher TRL, reflecting major recent investments in physical infrastructure.
Technology	Saudi Arabia	The system assigns a higher TRL, driven by accelerated pilot projects and strategic industry partnerships.
Data & Connectivity	Singapore	The system assigns a lower TRL due to its conservative assessment of full operational integration.
Societal & Market Readiness	Sweden	The system assigns a higher TRL, capturing strong, measurable public acceptance and market maturity indicators.

The system’s integrated assessment assigns Saudi Arabia an overall TRL of 7–8, indicating that AV technology has been demonstrated in operational environment. The primary strategic implication is the need to build upon current infrastructural and technological strengths while urgently addressing regulatory and societal gaps. Leveraging strong connectivity and strategic planning to develop robust regulations and foster public trust is essential for smooth AV integration, making policy development and public engagement critical priorities for future progress.

This research developed and validated a technology readiness assessment framework with LLM-enhanced document analysis that directly addresses the research questions.

In answer to RQ2, the framework was successfully developed as a hybrid system that combines structured expert logic with probabilistic LLM analysis under human guidance. To address RQ3, its reliability and robustness were rigorously evaluated using primary validation and statistical simulation. The system demonstrated high accuracy, with over 80% agreement with expert benchmarks and a low MAE of 0.44 TRL levels. The system shows broad usage through its consistent performance across 20 nations from different economic and geographic regions. The Monte Carlo simulation results show that the system maintained 92.1% stability when operating under unpredictable situations, while the confidence average reached 88.8%. Concerning RQ1, the application of this validated system enables a move from subjective assessments to a detailed evaluation of AV readiness. For Saudi Arabia, it provides policymakers with evidence-based tools for strategic decisions. The validated framework provides a consistent system for industry professionals to perform operational feasibility assessments. For the research community, it demonstrates the efficacy of a hybrid expert system for complex, multi-dimensional assessments and its potential applicability to other emerging technologies beyond AVs.

The validation of the readiness assessment framework provides concrete tools to advance Saudi Arabia's digital transformation under Vision 2030. The framework directly enables evidence-based policymaking, moving planning from uncertainty to a data-driven strategy. This reliable tool offers policymakers a clear, data-driven method to evaluate readiness and reduce investment risks, directly supporting the Vision's goals of technological leadership and a diversified digital economy. Planners and policymakers use the confirmed strengths in Technology, Infrastructure, and Data & Connectivity dimensions as a solid foundation for near-term pilot projects and initiatives. For Infrastructure, the strategy should build upon existing physical assets and ongoing 5G deployment, with an emphasis on modernizing legacy networks. For Data and Connectivity, it is essential to utilize current connectivity infrastructure while addressing cybersecurity and enhancing data interoperability. For Technology, the approach should leverage the established industrial roadmap and prioritize research and development commercialization. At the same time, the framework directs focused effort toward systematically closing the most critical gaps in regulatory frameworks and public engagement. For the Regulatory and Policy dimensions, the strategy should build on a mature national vision while strengthening the completeness of the framework, the scale of pilot projects, and policy codification. For the Societal and Market dimensions, the priority is to address the nascent stage of public outreach while systematically developing public acceptance, readiness growth, trust-building design, and economic viability through sustained engagement and targeted programs. Together, this integrated approach ensures that Saudi Arabia's journey toward leadership in smart mobility is built on a foundation that is both ambitious and practical.

6. Conclusions

This research addresses the challenge of assessing national readiness for AV integration. To bridge the gap between theoretical potential and practical deployment, a technology readiness assessment framework was developed. This research employed a mixed-methods approach to develop and validate the framework. The framework employs a novel hybrid architecture that integrates rule-based expert reasoning, a threshold-based scoring model for handling ambiguous edge cases, and an LLM-based document analysis to ensure both precision and adaptability. The system achieved 97.1% agreement with the published literature benchmarks, while the 92.1% stability rate validates its reliability. The application of this framework to Saudi Arabia provided key empirical findings. The nation operates at an overall TRL 7-8 level, with the strongest readiness in Infrastructure and Technology. However, the assessment crucially identified that Regulatory and Societal readiness present

significant constraints. This outcome underscores the framework's utility, transforming subjective evaluation into a structured, evidence-based tool suitable for guiding policy implementation and highlighting the urgent need for targeted public engagement and trust-building initiatives alongside continued technological development. This research is subject to several limitations that warrant consideration. A key limitation is the preliminary scope of the LLM module's validation. While its initial performance is encouraging, this component remains a proof-of-concept, and its robustness across a wider spectrum of real-world document types and complexities has not yet been fully established. The accuracy of the assessment is relay on the quality and availability of its data. While these sources provide valuable insights, they may not fully capture the most recent, non-public developments within industry and government. The expert interview sample was limited to seven participants. Although the participants represented multiple sectors relevant to AV deployment, the sample size limits statistical generalizability and may not fully reflect the diversity of stakeholder perspectives. Future research should expand participation and consider Delphi-based consensus methods involving a larger expert panel. Our research is focused on assessing the general national AV readiness ecosystem and does not attempt to model the distinct operational, regulatory, or infrastructural requirements unique to each deployment scenario such as (robotaxi/personal AV and public transport), which remain outside our scope. It is critical to note the defined scope of the framework. It is designed to assess technological and systemic readiness, not to evaluate overall program risk, financial viability, or detailed safety cases. As such, it should be integrated as a complementary tool within a broader ecosystem of risk management, cost-benefit analysis, and regulatory oversight for a fully comprehensive deployment strategy. Future research should build upon this foundation to enhance the framework's applicability, robustness, and strategic value. Future studies should compare different regions to reveal specific elements that influence AV adoption, thereby creating a comprehensive worldwide understanding of readiness barriers and potential benefits. The main technical approach focuses on developing the proof-of-concept LLM module into its final operational state. Building on its promising initial validation, future work will focus on enhancing its robustness by developing it into a fundamental self-updating feature that will operate on diverse complex documents from real-world environments. The system would operate independently to handle the changing literature, policies, and reports, which would keep its knowledge database up to date. To maximize the framework's utility for dynamic decision-making, future efforts should aim to integrate real-time data streams. Utilizing live data on infrastructure deployment, regulatory changes, and even public sentiment from social media would make the system more adaptive and responsive.

Author Contributions: Conceptualization, A.A.; Methodology, A.A. and S.E.; Software, A.A.; Validation, A.A.; Formal analysis, A.A.; Investigation, A.A.; Data curation, A.A.; Writing—original draft, A.A. and S.E.; Writing—review & editing, A.A., S.E. and M.M.; Visualization, A.A.; Supervision, S.E. and M.M.; Project administration, A.A., S.E. and M.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Deanship of Scientific Research (DSR) at King Abdulaziz University, Jeddah, under grant no. (IPP:1306-612-2026).

Institutional Review Board Statement: This study involved semi-structured interviews with experts and a public survey of adult participants. Participation was voluntary, informed consent was obtained from all participants, no sensitive personal data were collected, and all responses were anonymized prior to analysis. According to the applicable institutional requirements for minimal-risk research, formal ethical approval was not required.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study. Prior to participation, all participants were informed about the purpose of the study, the voluntary nature of their participation, their right to withdraw at any time, and the confidentiality of the information provided.

Data Availability Statement: The original contributions presented in this research are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: The authors, acknowledge with thanks DSR for technical and financial support. The authors are also grateful to the interviewees for their participation and contribution to this research.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Gracias, J.S.; Parnell, G.S.; Specking, E.; Pohl, E.A.; Buchanan, R. Smart Cities—A Structured Literature Review. *Smart Cities* **2023**, *6*, 1719–1743. <https://doi.org/10.3390/smartcities6040080>.
2. Bibri, S.E.; Alexandre, A.; Sharifi, A.; Krogstie, J. Environmentally Sustainable Smart Cities and Their Converging AI, IoT, and Big Data Technologies and Solutions: An Integrated Approach to an Extensive Literature Review. *Energy Inform.* **2023**, *6*, 9.
3. Attaran, H.; Kheibari, N.; Bahrepour, D. Toward Integrated Smart City: A New Model for Implementation and Design Challenges. *GeoJournal* **2022**, *87*, 511–526. <https://doi.org/10.1007/s10708-021-10560-w>.
4. Oladimeji, D.; Gupta, K.; Kose, N.A.; Gundogan, K.; Ge, L.; Liang, F. Smart Transportation: An Overview of Technologies and Applications. *Sensors* **2023**, *23*, 3880. <https://doi.org/10.3390/s23083880>.
5. Bishop, R. Intelligent Vehicle Applications Worldwide. *IEEE Intell. Syst.* **2000**, *15*, 78–81. <https://doi.org/10.1109/5254.820333>.
6. Song, X.; Gao, H.; Ding, T.; Gu, Y.; Liu, J.; Tian, K. A Review of the Motion Planning and Control Methods for Automated Vehicles. *Sensors* **2023**, *23*, 6140. <https://doi.org/10.3390/s23136140>.
7. Banke, J. *Technology Readiness Levels Demystified*; NASA: Washington, DC, USA, 2010.
8. Manning, C.G. *Technology Readiness Levels*; NASA: Washington, DC, USA, 2023.
9. Swaminathan, S.; Wishart, J.; Zhao, J.; Russo, B.; Rahimi, S. Adapting the Technology Readiness Level (TRL) Framework to Automated Vehicle Development. In *Proceedings of the WCX SAE World Congress Experience*; SAE International: Warrendale, PA, USA, 2025. <https://doi.org/10.4271/2025-01-8671>.
10. AlShizawi, H.O. The Future of Smart Transport Technologies: Examining Dubai’s Technological Readiness Levels (TRLs) to Adopt New Smart Transport Technologies. Master’s Thesis, Rochester Institute of Technology: Rochester, NY, USA, 2024.
11. Parasuraman, A.; Colby, C.L. An Updated and Streamlined Technology Readiness Index: TRI 2.0. *J. Serv. Res.* **2014**, *18*, 59–74. <https://doi.org/10.1177/1094670514539730>.
12. O’Hern, S.; St. Louis, R. Technology Readiness and Intentions to Use Conditionally Automated Vehicles. *Transp. Res. Part F Traffic Psychol. Behav.* **2023**, *94*, 1–8. <https://doi.org/10.1016/j.trf.2023.02.001>.
13. Department of Defense. *Technology Readiness Assessment Guidebook*; Department of Defense: Washington, DC, USA, 2023.
14. Mackin, M.; Bothwell, B.; Kaczmarek, S. *Technology Readiness Assessment Guide: Best Practices for Evaluating the Readiness of Technology for Use in Acquisition Programs and Projects*; Government Accountability Office (GAO): Washington, DC, USA, 2020.
15. Kimmel, W.M.; Beauchamp, P.M.; Frerking, M.A.; Kline, T.R.; Vassigh, K.K.; Willard, D.E.; Johnson, M.A.; Trenkle, T.G. *Technology Readiness Assessment Best Practices Guide*; National Aeronautics and Space Administration (NASA): Washington, DC, USA, 2020.
16. Threlfall, R. *Autonomous Vehicles Readiness Index*; KPMG International: London, UK, 2018.
17. Threlfall, R. *2019 Autonomous Vehicles Readiness Index*; KPMG International: London, UK, 2019.
18. Threlfall, R. *2020 Autonomous Vehicles Readiness Index*; KPMG International: London, UK, 2020.
19. Vangrattanachai, S. Infrastructure Readiness for Autonomous Vehicle Technology in Thailand. Ph.D. Thesis, Thammasat University, Bangkok, Thailand, 2023.
20. Gaitanidou, E.; Bekiaris, E. Consumer Acceptance in Measuring Greece’s Readiness for Transport Automation. *Future Transp.* **2022**, *2*, 644–658. <https://doi.org/10.3390/futuretransp2030035>.
21. Larsson, A. Are We Ready?: Swedish Road Operators Readiness for Automated Vehicle. Master’s Thesis, Uppsala University, Uppsala, Sweden, 2021.
22. Ho, J.S.; Sri Nusa Ahmad, L.Y.; Tan, B.C. A Conceptual Framework for Acceptance of Autonomous Vehicle in Malaysia. *Int. J. Manag. Financ. Account.* **2024**, *5*, 152–169. <https://doi.org/10.33093/ijomfa.2024.5.1.7>.
23. Sitinjak, C.; Said, M.H.M.; Hassim, J.Z.; Ali, H.M.; Khalid, R.M. Navigating the Future: Autonomous Vehicle Integration Challenges and Opportunities in Malaysia. *J. Infrastruct. Policy Dev.* **2024**, *8*, 2995. <https://doi.org/10.24294/jipd.v8i4.2995>.

24. Lim, W.; Lee, S.; Sung, B.; Cronin, S. Is Australia Ready for Autonomous Vehicles? Examining the Factors Influencing AV Adoption through Expert Interviews. *Australas. Mark. J.* **2024**, *32*, 308–322. <https://doi.org/10.1177/14413582231175152>.
25. Westerman, H.; Black, J. Preparing for Fully Autonomous Vehicles in Australian Cities: Land-Use Planning—Adapting, Transforming, and Innovating. *Sustainability* **2024**, *16*, 5492. <https://doi.org/10.3390/su16135492>.
26. Jiang, L.; Chen, H.; Chen, Z. City Readiness for Connected and Autonomous Vehicles: A Multi-Stakeholder and Multi-Criteria Analysis through Analytic Hierarchy Process. *Transp. Policy* **2022**, *128*, 13–24. <https://doi.org/10.1016/j.tranpol.2022.09.012>.
27. Salazar, G.; Russi-Vigoya, M.N. Technology readiness level as the foundation of human readiness level. *Ergon. Des.* **2021**, *29*, 25–29.
28. Alahmadi, A.; Elhag, S.; Meccawy, M. Assessing Public Acceptance of Autonomous Vehicles in Saudi Arabia. In *Proceedings of the 2025 IEEE/ACS 22nd International Conference on Computer Systems and Applications (AICCSA)*; IEEE: Piscataway, NJ, USA, 2025, pp. 1–7. <https://doi.org/10.1109/AICCSA66935.2025.11315380>.
29. Adeoye-Olatunde, O.A.; Olenik, N.L. Research and scholarly methods: Semi-structured interviews. *J. Am. Coll. Clin. Pharm.* **2021**, *4*, 1358–1367.
30. Khokhar, S.; Pathan, H.; Raheem, A.; Abbasi, A.M. Theory development in thematic analysis: Procedure and practice. *Rev. Appl. Manag. Soc. Sci.* **2020**, *3*, 423–433.
31. Fuchs, K. A systematic guide for conducting thematic analysis in qualitative tourism research. *J. Environ. Manag. Tour. (JEMT)* **2023**, *14*, 2696–2703.
32. Saunders, C.H.; Sierpe, A.; Von Plessen, C.; Kennedy, A.M.; Leviton, L.C.; Bernstein, S.L.; Goldwag, J.; King, J.R.; Marx, C.M.; Pogue, J.A.; et al. Practical thematic analysis: A guide for multidisciplinary health services research teams engaging in qualitative analysis. *Bmj* **2023**, *381*, e074256.
33. Kurse, T.K.; Gebresenbet, G.; Daba, G.F. Prospects for Implementation of Autonomous Vehicles and Associated Infrastructure in Developing Countries. *Infrastructures* **2024**, *9*, 237. <https://doi.org/10.3390/infrastructures9120237>.
34. Consulting, T. *Autonomous Vehicle Readiness Ranking 2025*; Tenet Consulting: Dubai, United Arab Emirates, 2025.
35. Paxton, P.; Curran, P.J.; Bollen, K.A.; Kirby, J.; Chen, F. Monte Carlo Experiments: Design and Implementation. *Struct. Equ. Model. A Multidiscip. J.* **2001**, *8*, 287–312. https://doi.org/10.1207/S15328007SEM0802_7.
36. Raychaudhuri, S. Introduction to monte carlo simulation. In *Proceedings of the 2008 Winter Simulation Conference*, Miami, FL, USA, 7–10 December 2008; pp. 91–100.
37. Shan, G. Monte Carlo Cross-Validation for a Study with Binary Outcome and Limited Sample Size. *BMC Med. Inform. Decis. Mak.* **2022**, *22*, 270. <https://doi.org/10.1186/s12911-022-02016-z>.
38. Koehler, E.; Brown, E.; Haneuse, S.J.P. On the Assessment of Monte Carlo Error in Simulation-Based Statistical Analyses. *Am. Stat.* **2009**, *63*, 155–162. <https://doi.org/10.1198/tast.2009.0030>.
39. Kroese, D.P.; Brereton, T.; Taimre, T.; Botev, Z.I. Why the Monte Carlo Method Is so Important Today. *Wiley Interdiscip. Rev. Comput. Stat.* **2014**, *6*, 386–392. <https://doi.org/10.1002/wics.1314>.
40. Mason, J.; Classen, S.; Wersal, J.; Sisiopiku, V. Construct Validity and Test–Retest Reliability of the Automated Vehicle User Perception Survey. *Front. Psychol.* **2021**, *12*, 626791. <https://doi.org/10.3389/fpsyg.2021.626791>.
41. Elangovan, N.; Sundaravel, E. Method of Preparing a Document for Survey Instrument Validation by Experts. *MethodsX* **2021**, *8*, 101326. <https://doi.org/10.1016/j.mex.2021.101326>.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.