

AI-POI-Sudan: Survey-Informed Design of a Random Forest Decision-Support Architecture for E-Government Service Quality

Elmouez Samir Abd Elhameed Mohamed¹, Anwer Hilal²

¹ Department of Computer Science, College of Postgraduate Studies, Sudan University of Science and Technology, Khartoum, Sudan

² Department of Information System, Faculty of Computer Science and Information Technology, Omdurman Islamic University, Omdurman, Sudan

ABSTRACT

Digital public-service evaluation must distinguish observed service deficiencies from predictions and the effects of proposed interventions. This study develops a survey-informed decision-support architecture for Sudanese e-government services, combining a descriptive empirical assessment with a specification for a Random Forest prediction module. Aggregate outputs from a doctoral survey of 216 respondents were examined, and seven service indicators with consistent item meanings were selected for analysis. Respondent counts, item means and standard deviations were retained, and differences from the Likert agreement reference of 4 were calculated as transparent diagnostic quantities. Four indicators had meanings below the neutral reference of 3: service availability (2.569), transactions without office visits (2.671), transaction speed (2.847), and regular information updates (2.926). Privacy-policy clarity (3.162), information completeness (3.088), and the absence of advanced technical-skill requirements (3.435) received higher mean ratings. Respondents outside Sudan accounted for 68.1% of the sample, restricting inference about conditions experienced by domestic users. The design contribution is a traceable architecture connecting survey indicators, a versioned feature dictionary, independently measured outcomes, a Random Forest module, and a scenario-support interface. The architecture separates descriptive gaps, predictive importance and intervention effects, and specifies validation and governance requirements. The study provides empirical service diagnostics and an implementable design specification; it does not establish predictive accuracy or causal policy effects. Its practical value lies in organizing service-improvement questions while defining the evidence required before operational adoption.

Keywords:— E-government service quality, Sudan, Random Forest, decision-support architecture, survey analysis, digital public services.

I. INTRODUCTION

E-government service quality is a question of how reliably public institutions enable people to obtain information and complete transactions, rather than simply whether an online portal exists. Established information-systems research distinguishes information, system and service quality, and relates these dimensions to user evaluations and use [1]. Service-quality traditions likewise emphasize the multidimensional nature of users' experiences [2]. In electronic settings, efficiency, availability, fulfillment and privacy provide complementary perspectives on performance [3], while e-GovQual adapts quality assessment to government services [4]. These foundations make service evaluation meaningful, but translating a set of survey results into a defensible improvement decision requires additional analytical and institutional work.

Empirical studies show why a single outcome or universal ranking is insufficient. Li and Shang analyze the relationship between e-government service quality, perceived value and continuous-use intention in China [5]. Pham et al. examine service quality, value, satisfaction and loyalty in Vietnam, reporting differentiated relationships among the quality dimensions [6]. Research on citizen attitudes and adoption also indicates that the use of digital government depends on

more than technical availability [7] [8]. Consequently, a decision-support system should preserve the meaning of its indicators and outcomes instead of treating all positive user evaluations as interchangeable measures of satisfaction.

The Sudanese setting makes this distinction particularly important. Abdallah and Fan develop an e-government assessment framework grounded in Sudan's institutional context [9], while Ahmed et al. investigate acceptance of Sudanese e-government services [10]. These studies motivate locally grounded assessment rather than automatic transfer of priorities from other countries. Users accessing a Sudanese service from abroad may also encounter different connectivity, service needs and institutional interfaces from users inside Sudan. An empirical assessment must therefore report whose experiences its data represent before recommending national interventions.

Artificial intelligence offers opportunities to support such assessment, but public-sector applications must be considered alongside institutional, technical and accountability challenges [11] [12]. Margetts similarly places AI within the broader problem of good governance [13]. A fitted model, an interactive dashboard and a justified public investment are distinct achievements. Treating them as equivalent can encourage claims that go beyond the evidence, especially

when service perceptions are measured once and institutional interventions have not been evaluated.

The gap addressed in this article is the insufficiently explicit connection between survey-based service diagnostics and the design requirements of a responsible decision-support system. Established quality models provide useful measurement foundations, but a service rating does not by itself specify a prediction task, justify an intervention, or assign institutional responsibility for a recommendation. AI-POI-Sudan addresses this translation problem by combining three elements: a descriptive assessment of selected survey indicators, an evidence-to-design traceability scheme, and a proposed Random Forest decision-support architecture with explicit validation and governance requirements. Its contribution lies in linking each diagnostic observation to a documented service meaning, a corresponding design requirement, and the additional evidence needed for evaluation. Following the distinction between description, explanation and prediction, the study presents a bounded empirical and design-oriented contribution; it does not claim a novel learning algorithm, demonstrated predictive superiority, or validated intervention effects [4].

The study addresses three questions: RQ1: What strengths and deficiencies are visible in the selected service indicators? RQ2: How does the sample composition constrain the interpretation of these diagnostics? RQ3: What architecture and validation requirements are needed to translate the evidence into responsible decision support?

II. THEORETICAL BACKGROUND AND RELATED WORK

A. Quality, User Outcomes and the Public-Service Context

The DeLone and McLean model provides a basis for distinguishing the quality of information, the functioning of a system and the support associated with its use [1]. SERVQUAL contributes a multidimensional service perspective [2]; E-S-QUAL adds characteristics of digitally mediated service delivery [3]; and e-GovQual focuses on government-service assessment [4]. These frameworks are complementary measurement resources. They should not be dismissed as incapable of useful analysis simply because they do not themselves specify a machine-learning deployment pipeline.

Quality dimensions and behavioral outcomes nevertheless require separate operational definitions. For example, a clear privacy policy is a service attribute, perceived value is an evaluation, satisfaction is an experiential outcome, and continued use is a behavioral intention. Pham et al. show that trustworthiness and fulfillment relate to perceived value in their Vietnamese sample [6]. Veeramootoo et al. find that satisfaction, system quality and habit are relevant to continued use of an e-filing service in Mauritius [15]. These results justify distinguishing constructs and service contexts rather

than equating one item's rating with every downstream outcome.

B. Trust, Privacy and Organizational Conditions

Trust provides an important bridge between quality assessment and the willingness to use digital government. Hooda et al. synthesize 67 studies and analyze trust within e-government success relationships [16]. This literature supports examining privacy-related perceptions, but it does not mean that privacy-policy clarity alone measures cybersecurity, trust in government or the probability of continued use. In the present study, the privacy item is interpreted narrowly as a reported evaluation of policy clarity, not an audit of technical safeguards.

Organizational conditions also influence the feasibility of improvement. The Sudan-focused studies by Abdallah and Fan and by Ahmed et al. situate evaluation within institutional and adoption concerns [9] [10]. In AI-POI-Sudan, the organizational perspective informs the design of data ownership, service responsibility, operational monitoring and human review. These are architectural requirements in this article, not newly validated latent constructs or measured causal mechanisms.

C. Random Forest and Responsible Predictive Interpretation

Random Forest averages the outputs of multiple decision trees and can represent nonlinear predictive relationships [17]. This makes it a candidate for a future survey-based prediction module, but choosing the algorithm does not establish the quality of the training data or guarantee generalization. Kapoor and Narayanan show why leakage can produce overly optimistic performance claims [18], and Cawley and Talbot demonstrate that model-selection practices can themselves create evaluation bias [19]. The architecture must therefore specify outcomes, feature availability and evaluation procedures before reporting predictive utility.

Model interpretation requires similar care. Strobl et al. demonstrate sources of bias in Random Forest variable-importance measures [20]. Local explanation approaches such as LIME illustrate a different objective: explaining an individual prediction rather than supplying a universal ordering of policy priorities [21]. Such methods may be useful extensions, but they are not presented as experiments performed here. The relevant design principle is that an explanation should remain tied to a specific model, output and input context.

D. From Empirical Diagnosis to an Information-System Artifact

Design-science research distinguishes the construction of an information-system artifact from evidence about its usefulness [22]. Peffers et al. articulate problem identification, solution objectives, development, demonstration, evaluation and communication as distinct activities [23]. This article uses those distinctions to organize a survey-informed architecture, without claiming a completed field evaluation of its utility.

Technical-debt research further emphasizes that data dependencies, configuration and maintenance are part of a deployed machine-learning system [24]. The NIST AI Risk Management Framework provides a complementary governance perspective through its Govern, Map, Measure and Manage functions [25].

III. RESEARCH METHODOLOGY

A. Study Design and Evidence Base

The article combines secondary descriptive analysis of aggregate outputs from the underlying doctoral survey with an evidence-informed architecture specification. The empirical unit is the respondent’s rating of a Sudanese e-government service attribute. The survey output reports 216 responses collected through Google Forms in April 2026. The present analysis uses the reported respondent frequencies and item-level means and standard deviations, not a new survey or a re-estimation from participant-level records. This distinction determines which analyses can be justified: descriptive comparisons are feasible, whereas subgroup effects, item correlations and inferential models cannot be reconstructed from these summaries alone.

The source research contains broader quality model and prototype documentation. For this article, the descriptive analysis is restricted to seven indicators whose item meanings can be traced consistently between the questionnaire wording and the reported descriptive summaries. The selection criterion is semantic traceability rather than statistical loading, predictive importance, or the magnitude of the observed rating. The indicators provide a bounded diagnostic view spanning perceived technical accessibility, transaction performance, service continuity, end-to-end completion, information quality, and privacy communication. Table I documents the selected indicators, their conceptual areas, reported means and standard deviations, and the rationale for their inclusion. These conceptual areas organize the present descriptive analysis; they are not newly estimated latent constructs, and the seven indicators are not presented as a validated psychometric scale.

TABLE I
TRACEABILITY OF THE SEVEN SELECTED
DIAGNOSTIC INDICATORS

Selected Indicator	Conceptual Area	Mean / SD	Inclusion Rationale
No advanced technical expertise needed	Perceived technical accessibility	3.435 / 1.102	Captures perceived requirements, distinguishing technical manageability skill

Selected Indicator	Conceptual Area	Mean / SD	Inclusion Rationale
Satisfactory transaction speed	Transaction performance	2.847 / 1.129	from transaction performance. Represents perceived timeliness of transaction completion.
Continuous service availability	Service continuity	2.569 / 1.127	Captures perceived continuity of access without attributing interruptions to a particular cause.
Transactions without office visits	End-to-end completion	2.671 / 1.153	Examines whether the digital channel supports completion without a physical visit.
Information completeness	Information quality: completeness	3.088 / 1.140	Represents the sufficiency of information needed to complete a service.
Regular information updates	Information quality: currency	2.926 / 1.159	Distinguishes the maintenance of current information from the presence of complete information.
Clarity of privacy policies	Privacy communication	3.162 / 1.196	Captures perceived policy clarity rather than technical security or overall trust.

Note: Means and standard deviations are retained from the reported doctoral survey outputs (N = 216). Indicators are identified by meaning rather than by assuming equivalence between numeric item identifiers across instrument versions. The conceptual areas and inclusion rationales describe their use in the present diagnostic analysis. Selection does not imply that these are the seven lowest-rated items, the most important predictors, or a comprehensive quality scale. Operational requirements and follow-up evidence are presented separately in Table IV.

B. Sample Description and Data Boundaries

Respondent characteristics are summarized by counts and percentages with denominator $N = 216$. Residence categories are retained as recorded: outside Sudan, Khartoum, other Sudanese towns or villages, and an unspecified “other” category. The unspecified category is not reassigned to a domestic or foreign group. Likewise, education and service-use categories are reported without inferring digital competence, citizenship status or service access from category membership. No population weights, synthetic respondent records, or post-hoc urban–rural balancing are applied to the aggregate outputs.

The online recruitment channel and achieved sample composition are treated as boundaries of interpretation. The article does not describe the sample as nationally representative, calculate a population response rate without a recruitment denominator, or infer a representative urban–rural distribution. The present analysis uses aggregate information only and includes no respondent identifiers; its ethical scope and data-availability boundaries are stated separately at the end of the article.

C. Descriptive Measures and Reference Comparisons

The selected items use a five-point agreement scale: 1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree and 5 = Strongly Agree. Item means and standard deviations are retained as reported to three decimal places. Means serve as compact descriptive summaries, not evidence that the ordinal response scale has been validated as an interval measure. Standard deviations describe variation in individual ratings and are not displayed as confidence intervals.

For transparent comparison, an item’s reference difference is defined as $g(j) = 4 - m(j)$, where $m(j)$ is its reported mean. The value 4 is the scale’s agreement anchor. It is not an empirically optimized policy target, a contractual service threshold or an expected intervention outcome. A larger positive difference means the average rating is farther below that reference. Rankings of these differences are therefore descriptive rankings of observed rating deficits, not rankings of predictive importance, cost-effectiveness or causal impact. No significance test is used to claim that adjacent means differ reliably.

Four descriptive checks were applied: respondent-category counts were checked against 216; percentages were recomputed from counts; selected-item summaries were cross-checked against the repeated analysis tables; and reference differences were calculated directly from the retained means. Where rounding affects percentage totals, the displayed value follows rounding to one decimal place. No respondent-level values were fabricated to reproduce aggregate statistics.

D. Evidence-to-Design Procedure

Design requirements were derived by linking the descriptive findings to the service activities they represent, then separating those findings from proposed analytical

capabilities. For example, a low availability rating motivates recording actual availability and failure events, but does not identify the cause of interruptions. A privacy-policy rating motivates a separate policy-clarity measure and governance review, not a declaration that a platform is secure. This traceability procedure follows the artifact-oriented emphasis on explicit objectives and evaluation in design-science research [22] [23].

The architectural specification identifies data inputs, a feature dictionary, independent outcomes, a Random Forest candidate module, and a dashboard with human review. It also defines a validation protocol. These are design outputs: no held-out prediction scores, model-comparison results, confidence intervals for forecasts, or measured improvements from interventions are claimed in this article. This separation allows the descriptive evidence to stand on its own without assigning untested capabilities to the architecture.

IV. EMPIRICAL FINDINGS

A. Respondent Profile and Its Interpretation

Table II reports the achieved sample. Men accounted for 146 responses (67.6%) and women for 70 (32.4%). A total of 197 respondents held a Bachelor’s degree, Master’s degree or PhD (91.2%). These figures characterize the respondents; they do not establish the educational or gender distribution of all users of Sudanese e-government services. Differences in who participates in an online survey can affect which service problems are visible in the resulting evidence.

TABLE II
RESPONDENT PROFILE (N = 216)

Characteristic / category	n	%
Gender		
Male	146	67.6
Female	70	32.4
Residence		
Outside Sudan	147	68.1
Khartoum	45	20.8
Other Sudan (town/village)	15	6.9
Others	9	4.2
Education		
High School	8	3.7
Bachelor's	113	52.3
Master's	50	23.1
PhD	34	15.7
Others	11	5.1

Characteristic / category	n	%
Service-use frequency		
Very seldom	65	30.1
Weekly	43	19.9
Monthly	39	18.1
Yearly	33	15.3
Biannual	19	8.8
Fortnightly	17	7.9

Note: Percentages are calculated from the recorded counts; totals may differ slightly from 100% because of rounding. “Others” in residence remains unspecified.

Residence is especially consequential. There were 147 respondents outside Sudan (68.1%), 45 in Khartoum (20.8%), 15 in other Sudanese towns or villages (6.9%), and nine in the unspecified category (4.2%). Thus 60 respondents were explicitly recorded in named domestic categories (27.8%). The remaining unspecified responses cannot be assumed to be domestic. The sample is therefore informative about a geographically mixed, diaspora-heavy respondent group, not a balanced assessment of conditions inside and outside Sudan.

The largest service-use category was “very seldom,” reported by 65 respondents (30.1%). Weekly users accounted for 43 respondents (19.9%). These categories indicate heterogeneous contact with digital public services. They do not explain why a person uses a service infrequently: infrequent need, channel preferences, service availability and other conditions could all contribute. Frequency is consequently used as a contextual descriptor rather than as proof of a service-adoption barrier.

B. Selected Service Ratings

Table III presents the selected indicators, with N = 216 for the reported analysis. The mean ratings range from 2.569 for continuous availability to 3.435 for not requiring advanced technical expertise. Four means fall below the neutral scale reference: availability, transactions without office visits, transaction speed, and regular information updates. The other three are above 3 but below the agreement reference of 4. None of these comparisons establishes the percentage of satisfied users or a difference that is statistically significant.

TABLE III
SELECTED SERVICE INDICATORS AND REFERENCE DIFFERENCES

Indicator	Mean	SD	4 – mean
No advanced technical expertise needed	3.435	1.102	0.565
Satisfactory transaction speed	2.847	1.129	1.153
Continuous service availability	2.569	1.127	1.431

Indicator	Mean	SD	4 – mean
Transactions without office visits	2.671	1.153	1.329
Information completeness	3.088	1.140	0.912
Regular information updates	2.926	1.159	1.074
Clarity of privacy policies	3.162	1.196	0.838

Note: Ratings range from 1 to 5. The difference from 4 is a descriptive comparison with the agreement anchor, not a predicted improvement. No significance ranking is implied.



Fig. 1. Mean ratings for the seven selected service indicators. The vertical references mark Neutral (3) and Agree (4); the figure shows neither predictive importance nor confidence intervals.

Continuous availability has the largest reference difference, 1.431 points, followed by transactions without office visits at 1.329 points. Transaction speed and information updating have differences of 1.153 and 1.074 points, respectively. These observations identify service tasks worth investigating: continuity, end-to-end completion, timeliness and the maintenance of service information. They do not determine whether a weakness originates in connectivity, institutional requirements, system design or another cause.

Privacy-policy clarity has a mean of 3.162 and a reference difference of 0.838, while information completeness has a mean of 3.088 and a difference of 0.912. Both are comparatively more favorable than the operational indicators, but neither reaches average agreement. The technical-skill item is the most favorable selected indicator, at 3.435. This pattern argues against reducing the diagnostic picture to a single universal concern: operational access, information and privacy communication can require different kinds of follow-up.

Standard deviations range from 1.102 to 1.196 across the selected items, indicating appreciable variation in reported experiences. Because participant-level covariances and group-specific responses are not analyzed here, the article does not attribute that variation to residence, gender, education or service type. Fig. 1 visualizes the means only; it should not be interpreted as an inferential comparison or a predictive ranking.

C. Translation into Design Requirements

Table IV connects the descriptive observations to system requirements and the additional operational evidence required to evaluate them. This is the bridge between the empirical and design components. The requirements preserve distinct questions: what respondents reported, what a service owner should investigate, what data the system would need, and what could eventually be evaluated. No requirement is treated as an intervention already implemented or an outcome already improved.

TABLE IV
TRACEABILITY FROM SURVEY FINDINGS TO DESIGN REQUIREMENTS

Observed pattern	Design requirement	Additional evidence
Availability and office-free completion below neutral	Separate access, continuity and completion indicators	Uptime, failures and visit requirements by service
Information updating below neutral	Version content and record update ownership	Content-update records and service requirement checks
Privacy clarity between neutral and agreement	Keep privacy communication distinct from security controls	Policy review and comprehension tasks
68.1% outside Sudan	Record access context; avoid national extrapolation	Domestic recruitment and subgroup data

Note: Requirements and follow-up measurements are design outputs, not implemented interventions or observed effects.

V. THE AI-POI-SUDAN DECISION-SUPPORT ARCHITECTURE

A. Architecture and Separation of Responsibilities

Fig. 2 specifies an architecture with separate data, modeling, inference and governance responsibilities. The design receives versioned survey measures and, where available, operational service records. A semantic register links each input to its wording, scale, date, service context, permitted use and transformation history. This register prevents a numeric column identifier from being treated as sufficient evidence of meaning. Perception measures remain distinct from event-based measures such as uptime or transaction duration; linking them requires a documented observation window and unit of analysis.

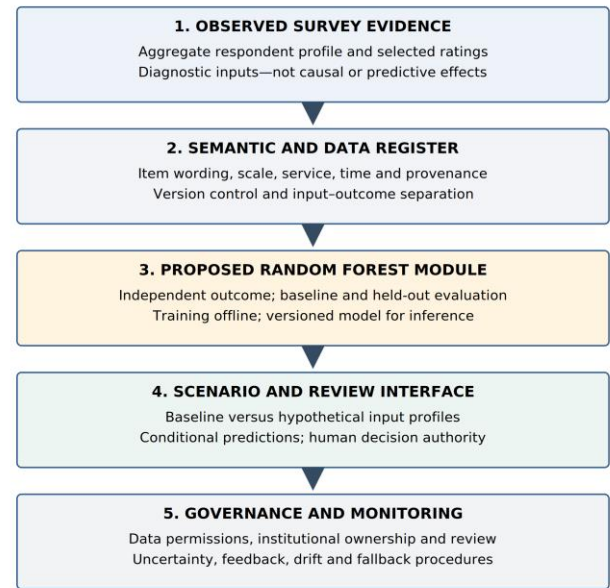


Fig. 2. Survey-informed AI-POI-Sudan architecture. The Random Forest, scenario and governance components are design specifications; the figure does not imply a validated deployed system.

Organizational responsibility is a cross-cutting component rather than an assumed effect estimated from the survey. A service owner is responsible for indicator definitions and maintenance; an analytical team is responsible for versioning and evaluation; an information-security function reviews permitted data flows; and the decision-maker retains responsibility for the interpretation and implementation of recommendations. This allocation reflects the institutional concerns emphasized in public-sector AI research [11] [12] and risk-management guidance [25].

B. Candidate Random Forest Module

A future prediction task must begin with a separately defined outcome. For example, a service-quality evaluation collected after a transaction could be predicted from attributes recorded beforehand, provided the outcome is not also present in the predictor set. An independently measured satisfaction scale would constitute a different prediction task from reconstructing a score calculated from its own inputs. The distinction between description and prediction is therefore operationalized in the data specification, not left to the naming of a model output [14] [18].

Random Forest is the candidate engine because it aggregates multiple regression trees [17]. Conceptually, its output is $f(x) = (1/T) \sum h(t)(x)$, where T is the number of trees and $h(t)$ is the prediction of tree t for the input profile x . This equation describes the proposed inference mechanism; it is not a newly trained model or a performance result. Training, preprocessing and hyperparameter selection should be separated from an untouched evaluation set. A simple baseline should be evaluated using the same outcome and splits before any advantage is claimed for the forest [19].

Performance reporting would include errors on the outcome's original scale, a clear description of the test population, and checks for stability across relevant user groups. Explanations should identify which model and prediction they describe. Global importance is not a probability of policy success, and local explanations do not replace a substantive account of the service process [20] [21]. The system should retain the option to withhold a prediction when inputs are incomplete, inconsistent or outside the model's intended scope.

C. Scenario Interface and Decision Outputs

The scenario component compares a baseline input profile x with a modified profile x^* . Its analytical output is the difference $f(x^*) - f(x)$. This is a conditional model comparison. It is not an estimate of what would happen if a government intervention caused the inputs to change. The interface should therefore distinguish observed values, user-entered hypothetical values and model-generated outputs, and should label scenarios in terms of assumptions rather than guaranteed benefits.

The proposed interface contains an indicator view, a data-quality and provenance view, a scenario comparison view, and a recommendation register. Each recommendation should state its evidence source, responsible service owner, required additional information and decision status. A rule-based suggestion should be identified as rule-based rather than portrayed as a learned recommendation. Financial costs, implementation times and legal feasibility must be supplied by the relevant institution; they cannot be inferred from a Likert-scale reference difference.

D. Validation and Deployment Requirements

Table V distinguishes the evaluation required for a future system from the evidence supplied by the present descriptive study. Predictive validation, explanation stability and usefulness to decision-makers are separate evaluation objectives. If forecast intervals are added, their empirical coverage must be assessed on appropriate held-out data. Variation across trees alone should not be presented as a calibrated confidence interval. Performance monitoring should also record changes in data definitions and service conditions, since these can alter model behavior after deployment [24].

TABLE V
PROPOSED VALIDATION AND GOVERNANCE PROTOCOL

Evaluation objective	Required procedure	Evidence to report
Outcome validity	Measure the outcome separately from inputs	Definition, timing and scoring
Predictive utility	Compare with a simple baseline on untouched data	MAE/RMSE and test population

Evaluation objective	Required procedure	Evidence to report
Interpretation stability	Check explanations across splits and correlated inputs	Rank variability and sensitivity
Decision-support utility	Use realistic tasks with intended users	Task errors, completion time and feedback
Operational governance	Assign ownership and review permissions and drift	Version history, approvals and fallback records

Note: These procedures specify future evaluation of the proposed architecture; they are not experiments claimed in this article.

A staged institutional pilot is more defensible than immediate nationwide deployment. An initial service-level implementation would establish data permissions, operational definitions, model responsibilities and fallback procedures. Subsequent use would depend on predictive evaluation and task-based assessment with intended users, not simply on the presence of a working interface. The architecture can support scheduled data exchange and controlled model updates, but no live government-database integration or completed user trial is asserted here. The proposed governance process follows the risk-based orientation of the NIST framework [25] and the institutional caution urged in governance research [13].

VI. DISCUSSION

A. Interpreting the Service Diagnostic Pattern

The empirical pattern suggests that the most evident concerns in the selected indicators relate to operational continuity and completing transactions without office visits. These are descriptive observations, not estimates of the causes of low satisfaction. They are compatible with the attention to availability and fulfillment in E-S-QUAL [3] and the multidimensional government-service perspective of e-GovQual [4]. The comparatively favorable technical-skill item also suggests that an interface can be perceived as manageable while the overall service journey remains constrained. Improving interface appearance alone would not necessarily address an observed difficulty in end-to-end completion, but the present analysis does not test that proposition.

The information indicators add another distinction. Completeness and regular updating are not identical capabilities. The first concerns whether necessary information is present; the second concerns whether it remains current. The mean for updating is below neutral, while the completeness mean is slightly above it. This is consistent with preserving information quality as a separate consideration in system-success research [1]. A practical next step would be to examine actual service descriptions, update records and conflicting requirements, rather than assume that all information deficiencies have the same solution.

B. Comparison with Previous Empirical Research

Li and Shang connect service quality, perceived value and continuous-use intention in China [5]. The present findings complement that account by identifying concrete diagnostic questions about service continuity and information maintenance. They do not replicate its structural relationships, because the present article reports descriptive outputs rather than an estimated chain model. Similarly, the Vietnam study by Pham et al. identifies trustworthiness and fulfillment as relevant to perceived value [6]; the Sudan-related diagnostics retain both operational and privacy-communication dimensions without imposing the Vietnamese model's relative effects on a different population.

Veeramootoo et al. examine an identifiable e-filing service and find continued use related to several factors, including satisfaction and habit [15]. Their service-specific design highlights a boundary of the present study: pooling general e-government experiences does not establish which attribute matters most for a tax transaction, licensing process or civil-record request. A service-specific deployment of AI-POI-Sudan would therefore require a service label, a clearly defined outcome and evidence collected for that service, rather than reusing one global ranking.

The trust literature provides context for privacy-policy clarity without turning it into a catch-all explanation. Hooda et al. demonstrate the relevance of trust within e-government success models [16]. Here, a privacy-clarity rating above neutral but below agreement indicates room for better communication, not proof of poor technical security or of security's causal dominance over usability. The distinction matters for intervention design: revising policy wording, changing access controls and increasing public trust are related but not equivalent activities.

Sudan-focused assessment and adoption research also supports a contextual interpretation [9] [10]. However, 68.1% of the current respondents were outside Sudan. Consequently, the observed ratings cannot be used as a direct measure of electricity or connectivity experienced by domestic users. Nor can one conclude that domestic constraints are unimportant merely because many respondents were connected from abroad. The survey provides one perspective on services associated with Sudan, and the architecture is designed to accommodate more explicit service and location information in subsequent use.

C. Contribution to Decision-Support Design

The methodological contribution is the separation of three quantities often conflated in policy-oriented analytics: an observed rating deficit, an item's predictive contribution, and the effect of implementing an intervention. The first can be calculated transparently from the descriptive evidence available here. The second requires an evaluated model with a defensible outcome. The third requires a suitable causal design and implementation information. This separation

applies Shmueli's distinction between analytical purposes [14] to the practical organization of a public-service dashboard.

The architectural contribution is the requirement that every input and recommendation be traceable to a service meaning, a data source and a responsible institutional actor. This extends the artifact-oriented emphasis on explicit objectives and evaluation [22] [23] into a setting where a visually persuasive interface should not be mistaken for validated decision support. The Random Forest component is consequently a proposed analytical capability within a broader governed system, not a substitute for measurement quality or a justification for automated resource allocation.

D. Practical Implications

The results support a manageable investigation agenda: examine service interruptions, identify where office visits remain necessary, review transaction delays, audit information-update practices, and test whether privacy explanations are understandable. These activities can be assigned to service owners and assessed with operational evidence. They do not require assuming that a survey rating already identifies the best investment or the largest financial return. The design also encourages phased adoption, human review and documented fallback procedures, consistent with research on public-sector AI challenges [11] [12] and responsible AI risk management [25].

VII. LIMITATIONS AND FUTURE RESEARCH

The empirical analysis is limited to aggregate outputs from one cross-sectional survey. Without participant-level records, the article cannot independently reconstruct response distributions, estimate covariances, assess subgroup differences or test measurement invariance. Means and standard deviations do not reveal how many respondents agreed with a statement, and the reference differences do not constitute validated dissatisfaction scores. The seven-indicator subset provides a bounded diagnostic view rather than comprehensive coverage of all quality, satisfaction and intention constructs.

The sample is predominantly male, highly educated and weighted toward respondents outside Sudan. The findings should not be generalized to all Sudanese citizens, rural users, people who cannot access digital channels, or users of a particular government service. The unspecified residence category was retained rather than assigned by inference. Future sampling should distinguish domestic and international access contexts and service types explicitly, with an appropriate recruitment and weighting strategy where population inference is intended.

The artifact contribution is an architecture specification, not a completed predictive-validation or public-sector implementation study. Future work should collect independently defined outcomes, develop a reproducible feature registry, compare Random Forest with appropriate

baselines, and evaluate the system on held-out observations and intended user tasks. Leakage checks and careful model selection are necessary safeguards rather than evidence already obtained [18] [19]. Monitoring, maintenance and organizational responsibilities must also be evaluated during deployment [24].

Finally, no measured policy intervention or cost-benefit study is reported. Scenario assumptions would need to be linked to practical implementation options, costs and timelines before they could support investment decisions. Longitudinal or quasi-experimental designs could investigate intervention effects, while qualitative feedback and service-specific case studies could clarify mechanisms behind the descriptive ratings. These extensions would build on the present diagnostic and architectural contribution rather than retrospectively convert it into causal evidence.

VIII. CONCLUSION

This study provides a bounded empirical assessment and a survey-informed decision-support design for Sudanese e-government services. Among seven traceable indicators from 216 respondents, continuous availability, transactions without office visits, transaction speed and information updating received mean ratings below the neutral scale reference. The geographically mixed sample, including 68.1% outside Sudan, defines the population of experience represented by those findings. The observations identify questions for service investigation; they do not establish causal priorities or quantify expected policy gains.

AI-POI-Sudan organizes the next analytical steps through a versioned feature dictionary, independently measured outcomes, a candidate Random Forest module, a scenario interface and institutional governance. Its contribution is to connect descriptive evidence with an explicit design and validation pathway while keeping observations, predictions and interventions distinct. Such a separation makes the framework more transparent and provides a practical foundation for subsequent service-specific empirical evaluation.

ETHICS STATEMENT

This article presents a secondary descriptive analysis of aggregate outputs from the underlying doctoral survey. No new participants were recruited for this article, and no respondent-level records or respondent identifiers were accessed or analyzed in the present analysis. The original thesis describes voluntary participation and confidentiality safeguards. However, the available thesis text does not establish a completed ethics-approval or exemption decision or document the actual consent procedure sufficiently to support a definitive statement about those procedures. No additional claim of ethics approval, exemption, or documented informed consent is therefore made here.

DATA AVAILABILITY STATEMENT

The aggregate respondent counts, item means, standard deviations, and reference differences used in the present descriptive analysis are reported in Tables II and III. The underlying survey instrument and additional aggregate summaries are documented in the doctoral thesis from which these outputs were drawn. No participant-level dataset or trained Random Forest model is provided with this article. The Random Forest component is an architectural specification rather than a newly trained and evaluated model. Access to underlying study records would require confirmation of the relevant institutional permissions and participant-consent conditions; no unrestricted access commitment is made in this article.

ACKNOWLEDGMENT

The authors express their sincere gratitude to the late Dr. Isabella M. Venter for her invaluable guidance, encouragement, and scholarly insight during the formative stages of this research. Although she passed away before the work was completed, her contributions helped shape its foundations. Her mentorship and academic influence are remembered with deep respect and appreciation.

REFERENCES

- [1] W. H. DeLone and E. R. McLean, "The DeLone and McLean model of information systems success: A ten-year update," *Journal of Management Information Systems*, vol. 19, no. 4, pp. 9–30, 2003, doi: 10.1080/07421222.2003.11045748.
- [2] A. Parasuraman, V. A. Zeithaml, and L. L. Berry, "SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality," *Journal of Retailing*, vol. 64, no. 1, pp. 12–40, 1988.
- [3] A. Parasuraman, V. A. Zeithaml, and A. Malhotra, "E-S-QUAL: A multiple-item scale for assessing electronic service quality," *Journal of Service Research*, vol. 7, no. 3, pp. 213–233, 2005, doi: 10.1177/1094670504271156.
- [4] X. Papadomichelaki and G. Mentzas, "e-GovQual: A multiple-item scale for assessing e-government service quality," *Government Information Quarterly*, vol. 29, no. 1, pp. 98–109, 2012, doi: 10.1016/j.giq.2011.08.011.
- [5] Y. Li and H. Shang, "Service quality, perceived value, and citizens' continuous-use intention regarding e-government: Empirical evidence from China," *Information & Management*, vol. 57, no. 3, Art. no. 103197, 2020, doi: 10.1016/j.im.2019.103197.
- [6] L. Pham, Y. B. Limbu, M. T. T. Le, and N. L. Nguyen, "E-government service quality, perceived value, satisfaction, and loyalty: Evidence from a newly emerging country," *Journal of Public Policy*, vol. 43, no. 4, pp. 812–833, 2023, doi: 10.1017/S0143814X23000296.
- [7] O. Al-Hujran, M. M. Al-Debei, A. Chatfield, and M. Migdadi, "The imperative of influencing citizen attitude toward e-government adoption and use," *Computers in*

- Human Behavior*, vol. 53, pp. 189–203, 2015, doi: 10.1016/j.chb.2015.06.025.
- [8] I. K. Mensah, “Factors influencing the intention of university students to adopt and use e-government services: An empirical evidence in China,” *SAGE Open*, vol. 9, no. 2, Art. no. 2158244019855823, 2019, doi: 10.1177/2158244019855823.
- [9] S. Abdallah and I. S. Fan, “Framework for e-government assessment in developing countries: Case study from Sudan,” *Electronic Government, an International Journal*, vol. 9, no. 2, pp. 158–177, 2012, doi: 10.1504/EG.2012.046267.
- [10] T. M. Ahmed, N. Alhadi, and M. E. Seliaman, “Acceptance of e-Government services in Sudan: An empirical investigation,” in Proc. 2015 International Conference on Cloud Computing (ICCC), 2015, pp. 1–4, doi: 10.1109/CLOUDCOMP.2015.7149625.
- [11] B. W. Wirtz, J. C. Weyerer, and C. Geyer, “Artificial intelligence and the public sector—Applications and challenges,” *International Journal of Public Administration*, vol. 42, no. 7, pp. 596–615, 2019, doi: 10.1080/01900692.2018.1498103.
- [12] T. Q. Sun and R. Medaglia, “Mapping the challenges of artificial intelligence in the public sector: Evidence from public healthcare,” *Government Information Quarterly*, vol. 36, no. 2, pp. 368–383, 2019, doi: 10.1016/j.giq.2018.09.008.
- [13] H. Margetts, “Rethinking AI for good governance,” *Daedalus*, vol. 151, no. 2, pp. 360–371, 2022, doi: 10.1162/daed_a_01922.
- [14] G. Shmueli, “To explain or to predict?,” *Statistical Science*, vol. 25, no. 3, pp. 289–310, 2010, doi: 10.1214/10-STS330.
- [15] N. Veeramootoo, R. Nunkoo, and Y. K. Dwivedi, “What determines success of an e-government service? Validation of an integrative model of e-filing continuance usage,” *Government Information Quarterly*, vol. 35, no. 2, pp. 161–174, 2018, doi: 10.1016/j.giq.2018.03.004.
- [16] A. Hooda et al., “Clarifying the role of e-government trust in e-government success models: A meta-analytic structural equation modeling approach,” *Australasian Journal of Information Systems*, vol. 27, 2023, doi: 10.3127/ajis.v27i0.4079.
- [17] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.
- [18] S. Kapoor and A. Narayanan, “Leakage and the reproducibility crisis in machine-learning-based science,” *Patterns*, vol. 4, no. 9, Art. no. 100804, 2023, doi: 10.1016/j.patter.2023.100804.
- [19] G. C. Cawley and N. L. C. Talbot, “On over-fitting in model selection and subsequent selection bias in performance evaluation,” *Journal of Machine Learning Research*, vol. 11, pp. 2079–2107, 2010.
- [20] C. Strobl, A.-L. Boulesteix, A. Zeileis, and T. Hothorn, “Bias in random forest variable importance measures: Illustrations, sources and a solution,” *BMC Bioinformatics*, vol. 8, Art. no. 25, 2007, doi: 10.1186/1471-2105-8-25.
- [21] M. T. Ribeiro, S. Singh, and C. Guestrin, ““Why should I trust you?”: Explaining the predictions of any classifier,” in Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 1135–1144, doi: 10.1145/2939672.2939778.
- [22] A. R. Hevner, S. T. March, J. Park, and S. Ram, “Design science in information systems research,” *MIS Quarterly*, vol. 28, no. 1, pp. 75–105, 2004, doi: 10.2307/25148625.
- [23] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, “A design science research methodology for information systems research,” *Journal of Management Information Systems*, vol. 24, no. 3, pp. 45–77, 2007, doi: 10.2753/MIS0742-1222240302.
- [24] D. Sculley et al., “Hidden technical debt in machine learning systems,” in *Advances in Neural Information Processing Systems*, vol. 28, 2015, pp. 2503–2511.
- [25] E. Tabassi, Artificial Intelligence Risk Management Framework (AI RMF 1.0), NIST AI 100-1. Gaithersburg, MD: National Institute of Standards and Technology, 2023, doi: 10.6028/NIST.AI.100-1.