Human-Centered Mobile Application Design for Saudi Digital Services under Vision 2030: Integrating Accessibility, Responsive Performance, Usability, and Unified Design Systems

Abstract:

Saudi Arabia's digital-service transformation increasingly depends on mobile applications that must work for diverse users, devices, network conditions, and service contexts. This review develops a human-centered framework for designing Saudi mobile services by integrating accessibility, responsive performance, usability, and unified design systems rather than treating these concerns as separate quality activities. Literature published from 2020 to 2025 was synthesized through a structured integrative review covering mobile usability, accessibility evaluation, non-functional testing, design-system practice, and Saudi digital-government adoption. The evidence shows that accessible interaction requires more than automated conformance checks; representative user testing, assistive-technology compatibility, clear Arabic and right-to-left behavior, and inclusive task design remain essential. Performance is similarly experiential: latency, unstable connectivity, device constraints, and visual reflow influence whether users perceive a service as understandable and trustworthy. Usability evidence supports combining task success, error analysis, standardized questionnaires, and behavioral feedback rather than relying on a single score. Design systems can institutionalize these requirements through reusable components, tokens, documentation, testing rules, and governance. The review proposes a closed-loop model in which measured user outcomes continuously refine shared components and service standards. For Vision 2030, the principal implication is that mobile quality should be governed as an integrated public-service capability, with accessibility, speed, usability, and consistency measured together.

Keywords: human-centered design; mobile applications; accessibility; responsive performance; usability; design systems; Saudi Arabia; Vision 2030; digital government

1. Introduction

Mobile applications have become a primary interface between institutions and citizens, changing the practical meaning of service accessibility. For Saudi digital services, the design problem is no longer limited to whether a transaction is technically available. A service must remain understandable, operable, efficient, and trustworthy across different abilities, languages, screen sizes, operating systems, device classes, and network conditions. Mobile usability research shows that context, efficiency, errors, and satisfaction interact [1,13]. Saudi evidence reinforces this point: assessments of mobile government interfaces have identified usability and accessibility weaknesses that can obstruct otherwise functional services [2], while evaluation of the Tawakkalna application found generally positive perceived usability alongside age-related difficulties that remained important for inclusive service delivery [10].

This creates a design challenge relevant to Vision 2030. Digital transformation can increase service reach, but adoption depends on more than deployment. Studies of Saudi e-government indicate that perceived usefulness, ease of use, trust, system quality, policy support, digital literacy, and security shape citizens' willingness to adopt and continue using digital channels [25–30]. Interface inconsistency, latency, inaccessible controls, and unstable layouts can increase abandonment, support demand, mistrust, and exclusion. Human-centered design reduces avoidable effort and makes service behavior predictable.

Four bodies of literature are especially relevant but are frequently treated separately. Accessibility research focuses on perceivability, operability, assistive technologies, and the participation of users with disabilities [3–8]. Usability research examines effectiveness, efficiency, learnability, errors, satisfaction, and measurement reliability [1,10,13–16]. Performance and testing studies address responsive layouts, non-functional requirements, device variability, and quality assurance [17–19]. Design-system research examines reusable components, consistency, documentation, governance, and standardization across products [20–22]. The central premise of this review is that Saudi digital services need these streams to be integrated. Accessibility without adequate performance can remain practically unusable; speed without clarity can accelerate errors; usability improvements without reusable standards can remain local; and a design system without evidence-based governance can simply reproduce defects at scale.

Accordingly, this paper synthesizes evidence from 2020–2025 to develop a practical framework for human-centered mobile application design in the Saudi context. The proposed model treats the application as a socio-technical service layer in which interface components, performance engineering, usability evidence, accessibility verification, organizational governance, and service outcomes are connected through continuous feedback. The review identifies decision principles and measurable controls adaptable across service domains. The objective is to support scalable quality while preserving the contextual requirements of Arabic interfaces, right-to-left interaction, inclusive access, and the growing expectation that public digital services should be seamless, reliable, and citizen-centered.

2. Aim, Objectives and Review Questions

The aim of this review is to develop an integrated decision-support framework for human-centered mobile application design in Saudi digital services under Vision 2030. The study deliberately moves beyond a conventional checklist approach by examining how four quality domains—accessibility, responsive performance, usability, and unified design systems—interact during design, delivery, measurement, and governance.

Five objectives guide the synthesis. First, the review identifies accessibility practices that reduce barriers for users with visual, motor, cognitive, and age-related needs, including the implications of Arabic and right-to-left interfaces. Second, it examines responsive performance as a user-experience variable shaped by latency, device capability, network instability, reflow, and interaction feedback. Third, it evaluates methods for measuring usability, with particular attention to task outcomes, standardized instruments, qualitative observation, and app-review evidence. Fourth, it investigates how design systems can convert isolated interface improvements into reusable institutional capabilities. Fifth, it translates these findings into a staged implementation model linked to inclusion, service completion, trust, consistency, and operational efficiency.

Three review questions follow: RQ1, which design and evaluation practices most reliably support accessible and usable mobile services? RQ2, how should performance, usability, and accessibility evidence be combined when governing interface quality? RQ3, how can a unified design system institutionalize these requirements across Saudi digital-service portfolios without suppressing service-specific needs?

3. Review Methodology

A structured integrative review was selected because the topic combines empirical human-computer interaction studies, software-quality research, accessibility evaluation, design-system practice, and Saudi digital-government evidence. Because the studies use heterogeneous outcomes, the method emphasized transparent selection, quality appraisal, thematic comparison, and conceptual synthesis rather than pooled effects.

Searches were designed around four concept blocks: mobile application or mobile government; accessibility or inclusive design; usability or user experience; and responsive performance, non-functional quality, or design systems. Saudi Arabia, Arabic, and right-to-left terms identified context-specific evidence. The core evidence window was 2020–2025. Peer-reviewed journal and conference studies in English were eligible when they addressed at least one focal design dimension and reported a method, evidence base, or evaluative framework that could inform mobile services. Saudi studies were prioritized for contextual claims, while international HCI and software-engineering studies were retained where the design mechanism was transferable.

The final evidence base was organized around thirty DOI-resolvable references. Relevance screening first removed sources whose focus was unrelated to mobile interaction, service quality, or institutional digital delivery. The remaining studies were appraised against four questions: whether the research design was explicit; whether the evaluated users, systems, or artifacts were relevant to the claim; whether the measures or analytical procedure were explained; and whether limitations permitted reasonable transfer to Saudi service contexts. The attached paper informed only broad structural organization; this manuscript uses an independent evidence base and search window.

Thematic coding proceeded in two passes. Deductive coding assigned findings to accessibility, responsive performance, usability, design systems, Saudi digital-service adoption, or cross-cutting governance. Inductive coding then captured recurring mechanisms such as assistive-technology compatibility, error prevention, feedback latency, device fragmentation, component reuse, measurement consistency, trust, and organizational ownership. Conflicting findings were retained. For example, standardized usability scores can support comparability, but observational evidence may reveal severe task barriers hidden by an acceptable average score [10,14–16]. Similarly, automated accessibility checks can improve coverage but cannot substitute for user testing with people who depend on assistive technologies [5,6].

The synthesis generated two outputs. The first is an evidence matrix linking each design dimension to risks, evaluation methods, and operational controls. The second is a closed-loop framework in which service telemetry and user research continuously update shared design components, accessibility rules, performance budgets, and usability standards. This approach is appropriate for a review paper because it converts heterogeneous evidence into testable design propositions without presenting invented primary data.

4. Saudi Digital-Service Context

Saudi Arabia is a distinctive setting because digital government has expanded alongside high smartphone use and rising expectations of seamless service. Saudi research demonstrates that adoption is influenced by more than technical availability. Sustainable e-government studies identify usability, security, performance, and service characteristics as relevant to citizens' willingness to use public platforms [27]. Technology-acceptance research similarly shows that usefulness, effort, awareness, self-efficacy, trust, and facilitating conditions shape adoption behavior [26,29]. More recent studies report that digital literacy, technical infrastructure, privacy, support services, and citizen trust continue to influence acceptance [25], while emerging-technology adoption inside government depends on strategic alignment, leadership, infrastructure readiness, skills, and interoperability [24].

Mobile delivery intensifies these issues because a single interface must serve users in different environments. A citizen may begin a task on a modern flagship phone over Wi-Fi and another may use an older device, enlarged text, a screen reader, or unstable mobile data. Accessibility and responsive performance therefore become public-service equity issues. Saudi evaluations have already shown that mobile government interfaces can contain accessibility and usability defects [2], and the evaluation of Tawakkalna indicates that overall usability can coexist with difficulties concentrated among particular age groups [10]. Aggregate satisfaction therefore cannot prove inclusion.

The Saudi context also introduces language and interaction requirements. Arabic interfaces require robust right-to-left layout behavior, mirrored navigation where appropriate, readable typography, correct number and date handling, and careful treatment of mixed Arabic-English content. These requirements should be implemented as reusable rules rather than repaired separately in each application. A unified design system can provide such institutional memory, but only if it is connected to accessibility testing, device testing, content standards, performance budgets, and governance. Vision 2030 alignment is therefore best understood not as adding a strategic label to an application, but as designing the operating capability required for inclusive, efficient, trusted digital service delivery.

5. Accessibility as a Service Quality Requirement

Accessibility is often reduced to conformance checking, yet mobile evidence shows that practical accessibility is broader. Paiva and colleagues' checklist work demonstrates the value of combining usability and accessibility criteria during mobile evaluation [3]. Research focused on visually impaired users shows that recurring barriers arise from unlabeled controls, unsuitable navigation structures, weak compatibility with screen readers, small targets, insufficient feedback, and interfaces that assume visual interpretation [5]. Studies of gesture typing and low-vision interaction further demonstrate that inclusive design can require interaction techniques that adapt to the behavior of users rather than merely exposing the same interface through assistive settings [7]. Users should be able to understand status, complete tasks, recover from errors, and confirm outcomes through different interaction modes.

Evaluation methodology matters. Pereira, Matos, and Duarte found that mobile accessibility assessment remains inconsistent and that user testing is underused despite its ability to reveal barriers that automated or manual standards checks may miss [6]. Di Gregorio similarly frames accessibility as a design problem requiring early interface decisions rather than late remediation [4]. For Saudi institutions, assurance should combine automated scanning, expert review, assistive-technology testing, and representative user sessions.

Arabic and right-to-left behavior require explicit attention within this model. Accessibility defects may arise when direction changes distort reading order, icons are mirrored incorrectly, numerals or mixed-language labels are announced confusingly, or enlarged text causes clipping. Information hierarchy and task meaning should remain equivalent across languages. Reusable bilingual components should therefore define semantic labels, minimum touch targets, dynamic type behavior, contrast states, focus patterns, and right-to-left rules as part of the component contract.

Age and cognitive load also matter. Nagro's evaluation of Tawakkalna found that perceived usability differed across age groups, reminding designers that public applications serve users with varied digital confidence [10]. Multi-step forms, dense terminology, timeouts, and hidden prerequisites can impose cognitive barriers despite technical conformance. Human-centered accessibility should consequently include plain language, progressive disclosure, clear progress indicators, recoverable sessions, explicit error prevention, and confirmation messages that state what happened and what the user should do next.

Accessibility governance should be measurable. Useful indicators include severe accessibility defects per release, percentage of critical journeys tested with assistive technologies, task completion by users with disabilities, unresolved semantic-label defects, text-resize pass rate, and defect recurrence across products. The most important governance change is to treat an accessibility finding as reusable evidence. If a defect occurs in a shared date picker, authentication field, modal, or navigation component, the correction should be made in the common component and propagated across services. This converts accessibility from repetitive audit work into continuous institutional learning.

Table 1. Evidence synthesis for integrated human-centered mobile service quality.

Dimension	**Evidence synthesis**	**Primary risk if isolated**	**Evaluation approach**	**Operational control**
Accessibility	Conformance must be paired with assistive-technology and representative-user evidence [3–8].	Formal compliance can coexist with blocked tasks.	Automated checks; expert review; screen-reader testing; inclusive task sessions.	Accessible component contracts; RTL rules; defect recurrence tracking.

Dimension	Evidence synthesis	Primary risk if isolated	Evaluation approach	Operational control
Responsive performance	Latency, instability, reflow and device variation affect task meaning and error risk [17–19].	Fast infrastructure can still produce slow or unstable interaction.	Device/network matrix; performance budgets; crash and latency telemetry.	Journey budgets; resilient states; layout-regression testing.
Usability	Effectiveness, efficiency, errors, satisfaction and context require mixed measurement [1,10–16].	Average scores can conceal severe journey or subgroup problems.	Task success; observation; standardized scales; support and review analysis.	Baseline journeys; release comparisons; root-cause feedback.
Unified design systems	Reusable components require governance, documentation and contribution practices [20–22].	Defects and inconsistency can scale across a portfolio.	Component QA; adoption analysis; version and contribution review.	Central foundations with federated evidence-based extensions.
Saudi digital-service context	Adoption depends on usability, trust, infrastructure, security and enabling conditions [23–30].	Deployment may not translate into inclusive or sustained use.	Segmented user research; acceptance measures; service analytics.	Portfolio governance aligned to inclusion, trust and completion.

6. Responsive Performance and Perceived Service Quality

Responsive performance is the second pillar because users experience time and stability as part of interface meaning. A screen that takes too long to confirm a payment, reloads unexpectedly, shifts controls during interaction, or becomes unresponsive under weak connectivity can cause errors even when the underlying transaction engine is correct. Research on non-functional testing in mobile applications shows that quality assurance must cover performance, compatibility, resource use, security, and related attributes across diverse devices and operating contexts [17]. Responsive-layout research likewise demonstrates the need to detect layout regressions across viewport conditions rather than assume that a design that works at one size will remain coherent elsewhere [18,19].

For human-centered mobile services, performance should therefore be governed through user-visible budgets, not only infrastructure metrics. Key measures include startup time, time to interactive state, response time after high-risk actions, percentage of sessions affected by crashes or freezes, layout-shift incidents, data transferred per journey, retry behavior, and success under constrained networks. The goal is predictable progress. Users should always receive immediate feedback that an action has been registered, distinguish pending from completed states, and avoid duplicate submission when the network is uncertain.

Device and network diversity are especially important for nationwide services. A design validated only on recent devices and fast connections can create hidden exclusion. Testing should therefore include representative lower-memory devices, multiple operating-system versions, enlarged text, Arabic and English, portrait and landscape where supported, and simulated bandwidth or latency constraints. The performance envelope of a service should be specified alongside its visual requirements.

Responsive behavior also intersects with accessibility. Large text may reflow screens and increase scroll distance; screen readers may expose delays that are less apparent visually; animated transitions can consume resources while distracting users; and lazy loading can disrupt focus if content appears unpredictably. Performance decisions should consequently be reviewed through accessibility and usability evidence rather than optimized in isolation. A practical pattern is to establish shared performance budgets for core components and critical journeys, then monitor real-world telemetry to identify regressions by device class, network type, application version, and language. This turns responsiveness into a continuously governed service-quality property.

7. Usability Measurement and Evidence

Usability determines whether users can achieve intended goals effectively, efficiently, and with acceptable effort. Weichbroth’s systematic review and later consolidated model show that mobile usability is multidimensional, with effectiveness, efficiency, satisfaction, learnability, errors, cognitive load, and contextual factors requiring coordinated assessment [1,13]. This argues against single-score assessment. Standardized questionnaires enable repeated measurement, but interpretation matters. Work on UMUX-LITE and SUS demonstrates that short measures can provide reliable perceived-usability signals and that wording and response-scale choices influence how instruments should be administered and compared [14–16].

Public-service evaluation should combine at least four forms of evidence. First, task metrics should measure completion, time, abandonment, retries, help-seeking, and critical errors for realistic journeys. Second, moderated or unmoderated observation should identify hesitation, misinterpretation, and workaround behavior. Third, standardized perceived-usability measures should support trend comparison across releases. Fourth, behavioral evidence from app reviews, support logs, analytics, and complaint categories should reveal issues at population scale. Nakamura and colleagues' review of app-store analysis shows that user experience is influenced by numerous factors and that review data can surface recurring concerns, although ratings alone do not establish causality [11]. Text-mining research in mobile health similarly demonstrates how large volumes of user commentary can expose experiential factors that formal testing may not capture [12].

Saudi evidence illustrates the value of mixed measurement. Tawakkalna achieved generally good perceived-usability scores, yet subgroup analysis still identified users who faced greater difficulty [10]. Averages can therefore conceal inequality. Mobile-government research should segment results by relevant characteristics such as age, assistive-technology use, language, digital confidence, device capability, and journey type where ethically and legally appropriate. The purpose is to reveal whether a design imposes systematically different burdens.

Usability testing must also focus on service outcomes rather than isolated screens. A visually clear form may still fail if prerequisites are discovered late, identity verification loops are confusing, an error message does not explain recovery, or users cannot distinguish submission from approval. End-to-end scenarios should include entry from notifications or deep links, authentication, consent, data entry, payment where relevant, document upload, status tracking, and recovery from interruption. Each critical journey needs an explicit success definition.

The strongest usability governance model is longitudinal. Baseline measures are established for core journeys; major changes are tested against the baseline; production analytics monitor completion and errors; qualitative research explains anomalies; and confirmed findings update shared components or content rules. Continuous measurement is preferable to one-off sign-off because mobile services evolve. It also creates a common evidence language between designers, developers, service owners, accessibility specialists, and senior decision-makers.

8. Unified Design Systems as Governance Infrastructure

A design system is more than a component library. Lamine and Cheng's empirical work shows that design-system practice involves coordination, contribution, governance, and support mechanisms that enable teams to maintain shared interface decisions [20]. Educational and practice-oriented research similarly emphasizes that design systems combine reusable assets with principles, documentation, and organizational processes [21]. In public services, the design system becomes governance infrastructure that encodes accessibility, responsive behavior, content, interaction, and evidence requirements.

Standardization should reduce repeated decisions without preventing contextual adaptation. Shared buttons, form fields, navigation, alerts, tables, cards, authentication patterns, consent components, file uploaders, date selectors, and status indicators can reduce visual drift and development effort. Reuse is beneficial only when components are well tested; defective components can multiply barriers across a portfolio. The system therefore needs versioning, ownership, acceptance criteria, deprecation rules, and product-team feedback.

Accessibility should be embedded at component level. Each component should document semantic behavior, focus order, keyboard or switch interaction where applicable, screen-reader labels, text scaling, contrast, error states, and Arabic/right-to-left behavior. Responsive requirements should specify allowable reflow, minimum supported widths, content wrapping, loading states, and performance expectations. Usability requirements should identify the intended task role and conditions where a component should not be used. This makes documentation an operational contract.

Paton and colleagues' work on HCI modelling and standardized interface components demonstrates the potential of shared design structures to improve consistency and make usability issues more visible [22]. For Saudi digital services, this approach can be extended through bilingual content tokens, standardized terminology, localized date and number patterns, and common service-status language. The system should also include analytics naming conventions so that comparable interactions produce comparable telemetry across applications.

Governance requires a contribution model. Central teams alone cannot anticipate every domain need, while unrestricted local changes erode consistency. A federated model is more suitable: a central design-system team owns foundational tokens, accessibility policy, quality gates, and core components; product teams can propose extensions supported by evidence; and changes with cross-service value are reviewed for inclusion in the shared system. Measures should include component adoption, defect recurrence, accessibility pass rate, reuse ratio, time saved in delivery, version fragmentation, and the proportion of high-volume journeys using approved patterns.

Unified design systems therefore connect strategic consistency with day-to-day engineering. They allow institutions to scale human-centered quality by moving knowledge from individual projects into a maintained product that serves multiple services.

9. Integrated Human-Centered Mobile Design Framework

The evidence supports a closed-loop framework with five connected layers. The first layer is user and service context: user goals, abilities, language, device, network, risk, and the administrative process being completed. The second layer is accessible interaction, where semantic structure, assistive-technology support, Arabic and right-to-left behavior, touch targets, text resizing, and error recovery are specified. The third layer is responsive performance, covering latency, stability, loading feedback, device compatibility, reflow, and constrained-network behavior. The fourth layer is usability evidence, combining task metrics, observation, standardized questionnaires, support data, and production analytics. The fifth layer is the unified design system, which converts validated patterns into reusable components, tokens, content rules, tests, and governance.

The sequence is not linear. Production outcomes feed back into earlier layers. A high abandonment rate may reveal a usability problem, but the cause could be slow identity verification, inaccessible focus behavior, unclear Arabic content, or a defective shared component. Investigation therefore requires cross-domain evidence. Once the cause is confirmed, remediation should occur at the highest reusable level: service process, shared content pattern, design-system component, performance standard, or application-specific implementation.

Table 1 summarizes the evidence relationships, while Figure 1 visualizes the closed loop. The contribution is integration. It rejects the common organizational pattern in which accessibility is audited near release, performance is owned only by engineers, usability is evaluated by a design team, and components are maintained separately. Instead, all four domains share measurable acceptance criteria and a common improvement cycle.

This structure can support procurement and assurance as well as internal delivery. A Saudi institution commissioning a mobile service can specify accessible task completion, performance budgets, usability thresholds, approved design-system use, and telemetry requirements as contractual quality outcomes. Vendors then have a clearer definition of “done,” and service owners can measure whether a release improves real user outcomes rather than merely meeting feature scope.

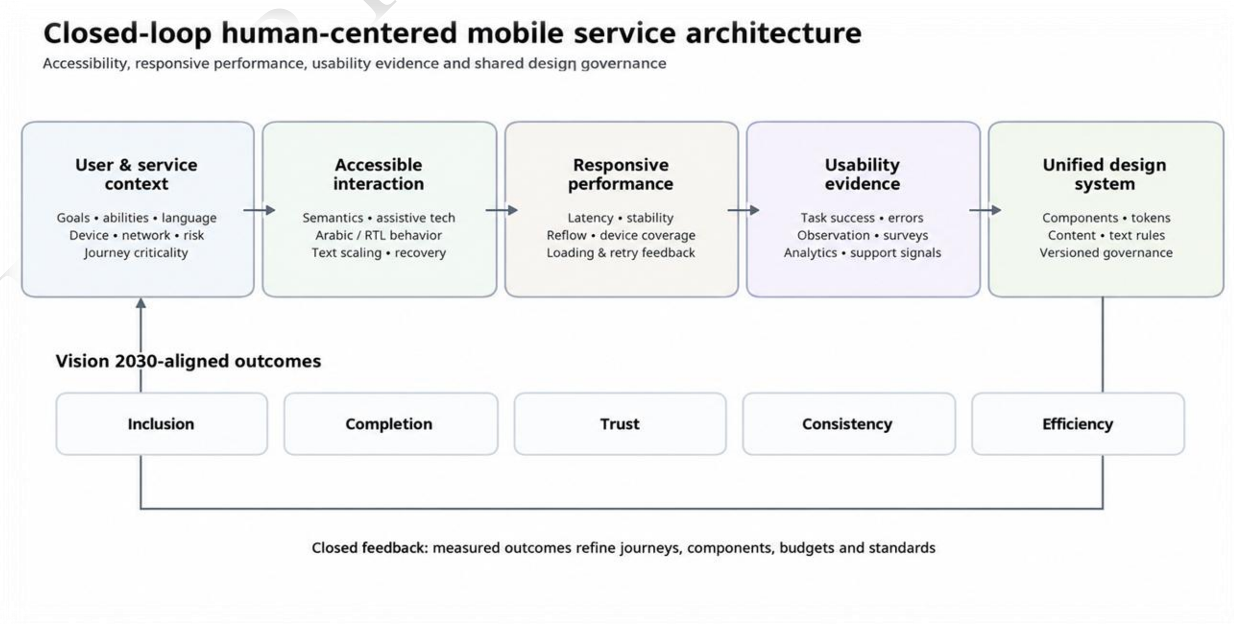

Figure 1. Closed-loop human-centered mobile service architecture for Saudi digital services.

10. Vision 2030 Implementation Roadmap

Implementation should be phased because institutions vary in design maturity, legacy constraints, product portfolios, and internal skills. The first stage is baseline and inventory. Teams identify critical mobile journeys, catalogue existing components, measure current usability and accessibility defects, profile performance by device and network, and document inconsistent interface patterns. This establishes a factual starting point.

The second stage builds accessible foundations. Institutions define minimum semantic, contrast, text-scaling, focus, touch-target, content, and right-to-left rules; establish assistive-technology testing; and prioritize remediation of high-volume or high-consequence journeys. The third stage integrates performance and usability. Critical journeys receive explicit performance budgets and task-success measures, with telemetry connected to research so that observed problems can be explained rather than merely counted. Dynamic testing of non-functional requirements supports this lifecycle approach [17].

The fourth stage operationalizes the unified design system. Frequently used validated patterns become centrally maintained components with bilingual documentation, automated checks, responsive behavior, and clear ownership. Product teams migrate progressively. The fifth stage establishes adaptive governance. A cross-functional design quality council reviews portfolio metrics, recurrent defects, new component proposals, and evidence from citizen research. High-severity problems trigger shared fixes and standards updates.

Figure 2 presents this maturity pathway. Table 2 links each stage to measurable indicators. Recommended outcome measures include accessible completion rate, critical-journey completion, median and tail response times, crash-free sessions, usability scores, severe error frequency, shared-component adoption, defect recurrence, support-contact rate, and user trust or satisfaction. Metrics should be interpreted together. A faster service with falling completion is not an improvement; a consistent interface that remains inaccessible is not mature; and high satisfaction among frequent users does not prove inclusion.

The Vision 2030 contribution lies in institutional capability. Human-centered mobile quality supports inclusion by reducing barriers, efficiency by reducing avoidable effort, trust by making service behavior predictable, and scalability by reusing verified patterns. It also strengthens localization because Arabic interaction rules and national service conventions become maintained assets rather than project-specific knowledge. A mature design system can therefore function as part of the digital public infrastructure: not a visible program in itself, but a shared mechanism that improves the quality of many citizen-facing services.

359 **Table 2. Maturity stages, evidence gates, and Vision 2030-aligned outcomes.**

Maturity stage	Priority actions	Key measures	Evidence gate	Vision 2030 contribution
1. Baseline & inventory	Map critical journeys, components, defects and device/network behavior.	Completion, severe defects, latency, support demand.	Documented baseline for high-volume journeys.	Visibility and prioritization.
2. Accessible foundations	Define semantic, text-scaling, assistive-tech and Arabic/RTL rules.	Assistive-tech pass rate; accessible completion; recurrence.	Critical journeys tested with representative access modes.	Inclusion and equitable access.
3. Performance + usability	Add performance budgets, task metrics and production telemetry.	Tail latency; crash-free sessions; task success; critical errors.	Regression against agreed journey baselines.	Reliable and efficient service delivery.
4. Unified design system	Move validated patterns into reusable components and bilingual guidance.	Approved-component adoption; reuse; version fragmentation.	Component-level accessibility, performance and usability acceptance.	Consistency and scalable localization.
5. Adaptive governance	Use portfolio metrics and shared defect learning to update standards.	Trust, satisfaction, recurrence, completion, support demand.	Cross-functional review and documented corrective action.	Sustained citizen-centered quality.

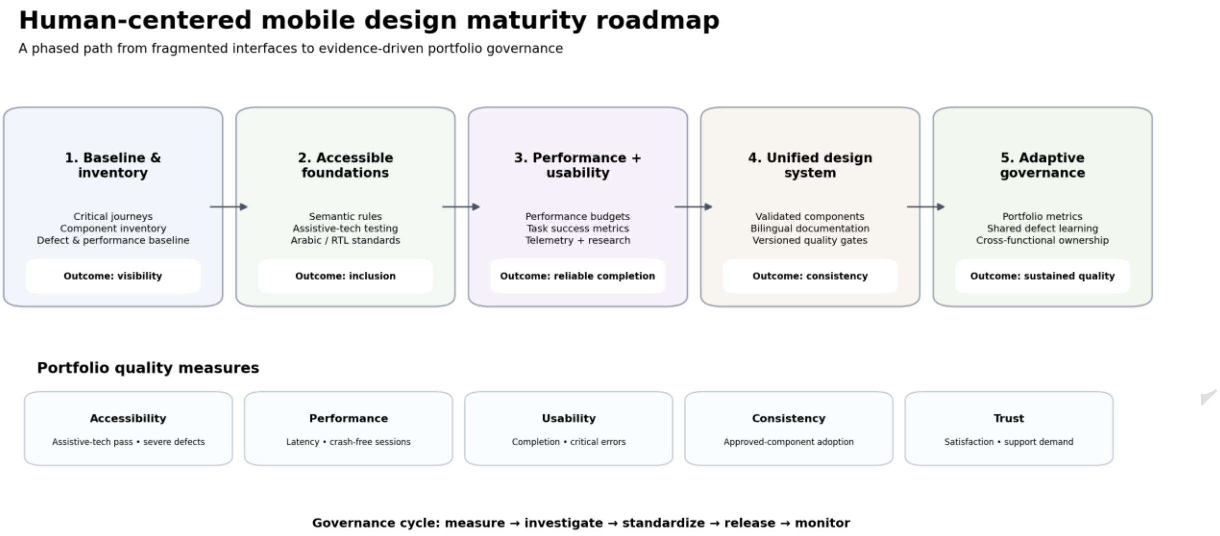

Figure 2. Human-centered mobile design maturity roadmap for Saudi digital-service portfolios.

11. Discussion, Research Gaps and Implications

The literature indicates strong progress in individual quality domains but less evidence on their integration. Accessibility studies increasingly recognize that automated checks are insufficient and that mobile evaluation needs representative user involvement [5,6]. Usability research offers mature measurement tools [14–16], while software-engineering studies provide methods for testing non-functional quality across mobile contexts [17–19]. Design-system research explains how shared practices can improve consistency and coordination [20,21]. Less developed is evidence on governing these mechanisms together across large digital-service portfolios.

This gap is relevant to Saudi Arabia. Existing studies provide useful snapshots of mobile government accessibility, application usability, e-government acceptance, sustainability, and emerging-technology adoption [2,10,24–30], but few studies connect interface-level quality to cross-agency design governance. Future work should test whether shared components reduce defects, performance budgets improve completion, Arabic/right-to-left standards reduce errors, and consistent patterns increase confidence.

Another gap concerns subgroup outcomes. Accessibility should not be inferred from overall satisfaction, and digital inclusion cannot be evaluated only through adoption counts. Research designs should include people with disabilities, older adults, users with low digital confidence, and varied device or network conditions. Longitudinal studies are also needed because mobile quality changes after release as operating systems, content, integrations, and user expectations evolve.

For practice, the main implication is organizational. Human-centered mobile quality requires shared ownership. Designers cannot solve infrastructure latency; engineers cannot infer all cognitive barriers from telemetry; accessibility specialists cannot repair inconsistent processes alone; and central design-system teams need evidence from real services. Institutions should therefore align product, engineering, content, accessibility, analytics, security, and service operations around common critical journeys and measurable outcomes. This integrated governance model is more demanding than a visual standard, but it is more likely to produce durable improvements.

12. Limitations and Future Research

This review has limitations. It synthesizes heterogeneous studies rather than conducting a meta-analysis, and international evidence may not transfer fully to Saudi administrative, linguistic, or cultural settings. The 2020–2025 window emphasizes recent practice but excludes useful earlier foundational HCI work. Different tools, instruments, devices, and domains also limit comparison.

Future research should test the proposed framework empirically across multiple Saudi applications. Useful designs include comparative studies before and after design-system adoption, controlled evaluation of Arabic/right-to-left components, accessibility testing with diverse assistive-technology users, performance studies on lower-end devices and constrained networks, and longitudinal analysis linking interface changes to task completion, support demand, and trust. Research should also examine governance: contribution models, component ownership, release processes, and the economic value of preventing repeated defects across portfolios.

13. Conclusion

Human-centered mobile design for Saudi digital services requires accessibility, performance, usability, and consistency to be treated as one service-quality system. The reviewed evidence shows that mobile accessibility depends on meaningful task completion with assistive technologies and diverse users; responsive performance depends on predictable interaction under realistic device and network conditions; usability requires multiple forms of evidence; and design systems can convert validated decisions into reusable institutional assets.

The proposed framework connects these domains through continuous feedback. User outcomes are measured in production and research, causes are investigated across interaction and technical layers, and validated improvements are incorporated into shared components, content rules, performance standards, and governance. This reduces the risk that the same barrier will be rediscovered separately in multiple applications.

For Vision 2030, the strategic value is practical. Inclusive and reliable mobile services can lower citizen effort, strengthen trust, improve completion, and increase the scalability of digital transformation. The priority should shift from project-specific quality toward a maintained capability for human-centered mobile service design. Such a capability is measurable, reusable, bilingual, evidence-driven, and adaptable, allowing Saudi institutions to improve digital services continuously as technologies, user expectations, and service responsibilities evolve.

References

1. Weichbroth, P. Usability of Mobile Applications: A Systematic Literature Study. IEEE Access 2020, 8, 55563–55577. https://doi.org/10.1109/ACCESS.2020.2981892

2. Al-Sakran, H.O.; Alsudairi, M.A. Usability and Accessibility Assessment of Saudi Arabia Mobile E-Government Websites. IEEE Access 2021, 9, 48254–48275. https://doi.org/10.1109/ACCESS.2021.3068917

3. Paiva, D.M.B.; Batista, M.H.S.; Zaina, L.A.M.; Fortes, R.P.M. Acc-MobileCheck: a Checklist for Usability and Accessibility Evaluation of Mobile Applications. CLEI Electronic Journal 2020, 23(2). https://doi.org/10.19153/cleiej.23.2.3

4. Di Gregorio, M. Accessible Applications—Study and Design of User Interfaces to Support Users with Disabilities. Proceedings of the 2021 International Conference on Multimodal Interaction 2021, 3 pp. https://doi.org/10.1145/3462244.3481281

5. Al-Razgan, M.; Almoaiqel, S.; Alrajhi, N.; Alhumegani, A.; Alshehri, A.; Alnefaie, B.; AlKhamiss, R.; Rushdi, S. A systematic literature review on the usability of mobile applications for visually impaired users. PeerJ Computer Science 2021, 7, e771. https://doi.org/10.7717/peerj-cs.771

6. Pereira, L.S.; Matos, M.; Duarte, C. Exploring Mobile Device Accessibility: Challenges, Insights, and Recommendations for Evaluation Methodologies. Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems 2024, Article 964. https://doi.org/10.1145/3613904.3642526

7. Zhang, D.; Seiple, W.H.; Li, Z.; Ramakrishnan, I.V.; Ashok, V.; Bi, X. Accessible Gesture Typing on Smartphones for People with Low Vision. Proceedings of the ACM Symposium on User Interface Software and Technology 2024, Article 86. https://doi.org/10.1145/3654777.3676447

8. Alayed, A. Are Saudi Arabian banks' mobile applications accessible for blind or partially sighted users?: a customers' perspective and evaluation. Universal Access in the Information Society 2025, 24, 585–606. https://doi.org/10.1007/s10209-023-01082-y

9. Alabdulrahman, R.; Alqahtani, F.; Binkhonain, M. Uncovering Usability Issues in Mobile Government Applications: A User-Centered Approach. IEEE Access 2025, 13, 169352–169366. https://doi.org/10.1109/ACCESS.2025.3614165

10. Nagro, S.A. User Experiences and Usability Evaluation of COVID-19 Application. Journal of Computer Science 2023, 19(3), 372–388. https://doi.org/10.3844/jcssp.2023.372.388

11. Nakamura, W.T.; de Oliveira, E.C.C.; de Oliveira, E.H.T.; Redmiles, D.F.; Conte, T. What factors affect the UX in mobile apps? A systematic mapping study on the analysis of app store reviews. Journal of Systems and Software 2022, 193, 111462. https://doi.org/10.1016/j.jss.2022.111462

12. Pal, S.; Biswas, B.; Gupta, R.; Kumar, A.; Gupta, S. Exploring the factors that affect user experience in mobile-health applications: A text-mining and machine-learning approach. Journal of Business Research 2023, 156, 113484. https://doi.org/10.1016/j.jbusres.2022.113484

13. Weichbroth, P. Usability of Mobile Applications: A Consolidated Model. IEEE Access 2024, 12, 62378–62391. https://doi.org/10.1109/ACCESS.2024.3395528

14. Borsci, S.; Buckle, P.; Walne, S. Is the LITE version of the usability metric for user experience (UMUX-LITE) a reliable tool to support rapid assessment of new healthcare technology? Applied Ergonomics 2020, 84, 103007. https://doi.org/10.1016/j.apergo.2019.103007

15. Lewis, J.R. Measuring User Experience With 3, 5, 7, or 11 Points: Does It Matter? Human Factors 2021, 63(6), 999–1011. https://doi.org/10.1177/0018720819881312

16. Kortum, P.; Acemyan, C.Z.; Oswald, F.L. Is It Time to Go Positive? Assessing the Positively Worded System Usability Scale (SUS). Human Factors 2021, 63(6), 987–998. https://doi.org/10.1177/0018720819881556

17. Júnior, M.C.; Amalfitano, D.; Garcés, L.; Fasolino, A.R.; Andrade, S.A.; Delamaro, M. Dynamic Testing Techniques of Non-functional Requirements in Mobile Apps: A Systematic Mapping Study. ACM Computing Surveys 2022, 54(10s), Article 214. https://doi.org/10.1145/3507903

18. Walsh, T.A.; Kapfhammer, G.M.; McMinn, P. Automatically identifying potential regressions in the layout of responsive web pages. Software Testing, Verification and Reliability 2020, 30, e1748. https://doi.org/10.1002/stvr.1748

19. Kumar, B.A. Mobile-First Usability Guideline for Responsive E-Commerce Websites. International Journal of Web Portals 2022, 14(1), 1–12. https://doi.org/10.4018/IJWP.310060

20. Lamine, Y.; Cheng, J. Understanding and supporting the design systems practice. Empirical Software Engineering 2022, 27(6), 146. https://doi.org/10.1007/s10664-022-10181-y

21. Rose, E.J.; MacDonald, C.M.; Putnam, C. Design Systems: A scalable model for teaching design systems for UX. Proceedings of the 5th Annual Symposium on HCI Education 2023, 5–7. https://doi.org/10.1145/3587399.3587403

22. Paton, C.; Borycki, E.M.; Warren, J.; Kushniruk, A.W.; English, M. HCI-modelling for improving the clinical usability of digital health technologies. Methods 2024, 227, 60–77. https://doi.org/10.1016/j.ymeth.2024.04.019

23. Nusir, M. Enhancing M-Government service adoption in Saudi Arabia: the role of mobile edge computing in moderating technology acceptance. Transforming Government: People, Process and Policy 2025, 19(1), 130–158. https://doi.org/10.1108/TG-07-2024-0176

24. Alshahrani, A. Adopting Emerging Technologies in Digital Government: A Multi-Case Analysis of Drivers, Enablers, and Challenges in Saudi Arabia. Digital Government: Research and Practice 2025, 6(3), Article 36. https://doi.org/10.1145/3719297

25. Hasan, A.; Alenazy, A.A.; Habib, S.; Husain, S. Examining the drivers and barriers to adoption of e-government services in Saudi Arabia. Journal of Innovative Digital Transformation 2024, 1(2), 139–157. https://doi.org/10.1108/JIDT-09-2023-0019

26. Almamy, A.S.M. Understanding Factors Affecting E-Government Adoption in Saudi Arabia: The Role of Religiosity. International Journal of Customer Relationship Marketing and Management 2022, 13(1), 1–15. https://doi.org/10.4018/IJCRMM.289209

27. Aljarallah, S.; Lock, R. An Investigation into Sustainable e-Government in Saudi Arabia. Electronic Journal of e-Government 2020, 18(1), 1–16. https://doi.org/10.34190/EJEG.18.1.001

28. Kala, D.; Chaubey, D.S.; Meet, R.K.; Al-Adwan, A.S. Impact of User Satisfaction With E-Government Services on Continuance Use Intention and Citizen Trust Using TAM-ISSM Framework. Interdisciplinary Journal of Information, Knowledge, and Management 2024, 19, Article 1. https://doi.org/10.28945/5248

29. Al-Rahmi, W.M.; Uddin, M.; Alkhalaf, S.; Al-Dhlan, K.A.; Cifuentes-Faura, J.; Al-Rahmi, A.M.; Al-Adwan, A.S. Validation of an Integrated IS Success Model in the Study of E-Government. Mobile Information Systems 2022, 2022, 8909724. https://doi.org/10.1155/2022/8909724

30. Al-Zahrani, M.S. Integrating IS success model with cybersecurity factors for e-government implementation in the Kingdom of Saudi Arabia. International Journal of Electrical and Computer Engineering 2020, 10(5), 4937–4955. https://doi.org/10.11591/ijece.v10i5.pp4937-4955