User Experience Evaluation in Government Applications: A Systematic Review

Emmy Hossain[1], Noris Mohd Norowi[2], Azrina Kamaruddin[3], Hazura Zulzalil[4]
Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia [1, 2, 3, 4]
Faculty of Computer Science and Information Technology, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia[1]

***Abstract*—Evaluating the User Experience (UX) of government applications is becoming increasingly crucial as governments deploy public services online. Nevertheless, research in this area remains fragmented. Correspondingly, this study presents a systematic review of UX evaluation in government applications to address the following Research Questions (RQs): What UX evaluation approaches and UX dimensions have been employed in the UX evaluation of government applications, and how do domains, contextual, and cultural considerations influence the UX evaluation of government applications? Kitchenham and Charters' guidelines, as well as the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA), are employed to guide this review. Moreover, recent studies from Scopus and Web of Science (WoS) databases between the years 2023 and 2025 were retrieved using a predefined review protocol. After applying the inclusion and exclusion criteria and subjecting the studies to quality assessment, the final number of retained studies for this review is 19. The analysis reveals four key themes: diversity in UX evaluation approaches, the range of UX dimensions evaluated, the range of domains evaluated, and the contextual and cultural considerations in UX evaluations. The findings reveal that UX evaluations of government applications are predominantly usability-focused, while hedonic, emotional, and cultural dimensions receive limited and inconsistent attention. In addition, the review highlights that UX evaluations for government applications should encompass both technical and pragmatic aspects, as well as domain-specific, cultural, and contextual dimensions. Accordingly, strengthening these evaluations can lead to more inclusive and meaningful assessments, resulting in government applications that offer better UX. Overall, the findings of this review may serve as a reference for future work and advance the field of UX evaluation, especially in the context of government applications.**

Keywords*—User Experience Evaluation; UX evaluation; government applications; e-government; systematic review

I. Introduction

Within the domain of human-computer interaction, User Experience (UX) has emerged as a fundamental aspect in the quality of interaction between users and digital systems, extending beyond traditional usability. In [1], UX is defined as encompassing all aspects of users' interaction with products and services, including perceptions, emotions, and responses that arise before, during, and after use [2], [3]. Similar to other forms of computing technology, all applications, including those used by governments, have UX as an integral component. Thus, it is crucial for UX to be considered during the design, development, and evaluation stages of these applications. This is to ensure a positive UX when interacting with, and/or anticipating interaction with, the applications. Consequently, to benefit from what the government intended to provide and share through its applications. Note that poor UX can cause repercussions, such as having difficulties in using the application. This includes missing out on crucial information, frustration, and abandonment, which can defeat the purpose of having these e-government initiatives [4].

In the context of interaction design, the lifecycle of producing applications (design, development, and evaluation stages) includes evaluation, a core component that is centered on both usability and UX [5]. In particular, UX evaluation is the process of collecting and analyzing data about users' or potential users' UX when interacting with artifacts. Examples of artifacts include screen sketches, prototypes, system components, a specific function, the entire workflow, the complete system or application, and safety features, to name a few. Therefore, it is crucial to conduct an evaluation for the following reasons: a rigorous evaluation helps enhance overall quality, allows users to understand their expectations, and ensures the product aligns with their needs and preferences. This guarantees that products are well-designed, as good experiences lead to desirable outcomes such as adoption, retention, use, sales, downloads, and recommendations. It also enables the identification and rectification of any issues before the product is released to the market [4], [6].

Since the early 1990s, the field of Information and Communication Technologies (ICTs) has undergone rapid progress, resulting in significant improvements across various societal areas, including, but not limited to, agriculture, communication, education, industry, transportation, government, and everyday life. In ICTs for government, one of the most notable innovations to emerge is electronic government (e-government). Notably, e-government is an umbrella term for the use of ICTs to access and deliver government services, disseminate government information, communicate between the government and its citizens, agencies, and stakeholders, and support internal government operations. The aim of e-government is to enhance the efficiency and effectiveness of government operations and accountability, as well as improve the government's engagement with citizens [7], [8]. Furthermore, to keep pace with technological advancements and provide better, faster, and more convenient services to citizens, governments offer applications through both mobile and web platforms, enabling citizens and stakeholders to access government services and engage with the government. As the

use of these applications and the reliance of the public on government applications grow, these applications must be designed and evaluated to support meaningful engagement and provide a positive UX.

Therefore, evaluating UX in the context of government applications is a core necessity, as UX is progressively acknowledged as essential for products and services, including those in the government sector [3]. While governments have increasingly delivered services through government applications, and the usage of these applications has risen, studies on their UX evaluation remain fragmented and dispersed, with varied approaches and focuses [9], [10], [11]. In [9], the authors showed that the studies addressing UX factors such as usability, trust and satisfaction are inconsistent across different service types and regions. The works of [10] and [11] demonstrated the common usage of usability evaluation approaches such as System Usability Scale (SUS), Smartphone Usability Questionnaire (SURE), heuristic evaluation and usability inquiry, to conduct UX evaluation for e-governments applications. While usability is a core component of UX, conducting UX evaluation using usability evaluation approaches is insufficient, with only usability as the focal point and not including other aspects of the user experience [37]. Additionally, the development of cumulative knowledge that can inform better UX evaluation is hindered, as is the potential for comparability across studies in the field. Meanwhile, several UX evaluation works have incorporated a broader range of UX dimensions beyond usability, and adopted evaluation approaches other than usability evaluation, such as [2] in online systems, [38] in AI-powered systems, [39] in business intelligence visualization, [40] in blockchain, and [41] in education. However, these works are domain-specific, designed with consideration of the domain's unique UX dimensions, and hence these UX evaluations are not directly transferable to the government application context. In contrast to the existing related studies, this study offers a systematic review that identifies and synthesizes UX evaluation approaches, dimensions, domains, and contextual and cultural considerations reported in the literature on government applications. Hence, this systematic review is warranted to investigate the existing literature on approaches, dimensions, domains, and considerations, contextualized and culturally. This ensures clearer directions for future research and practice in the UX evaluation of government applications.

Although there are several review studies that are related to this study, the existing reviews demonstrate notable limitations that motivate the present study. First, some review studies on government applications concentrated on usability evaluation rather than UX evaluation, leading to a narrower evaluative scope [10, 11]. Second, in [3], the authors examined the UX dimensions of e-procurement systems, which are only a subset of government applications and do not fully represent the larger domain of government applications. Other reviews employed a general perspective without focusing on domains: [42] reviewed UX evaluation techniques, factors influencing UX, and benefits of evaluating UX without anchoring their study in any specific domain. Another study that used the general perspective is [43], which focused on reviewing methods and metrics of UX evaluation. Lastly, [44]'s review study focused on the evaluation of public service quality instead of the evaluation of its applications (e.g., websites and mobile applications), and did not explicitly include UX evaluation in their study. These limitations, when viewed collectively, show that existing reviews do not provide a recent, domain-specific account of UX evaluation in government applications, with contextual and cultural considerations taken into account. Notably, the earlier reviews have not systematically examined how domain and cultural and contextual considerations are reflected in UX evaluations in government applications. To the best of the authors' knowledge, this is the first review that consolidated UX evaluation approaches and UX dimensions in government applications with the inclusion of domain-specificity, and contextual and cultural considerations. Also, concentrating this review on literature published between 2023 and 2025 enables the capture of recent developments in UX evaluation in government applications that reflect shifts in electronic government strategies, public service delivery via mobile applications, emerging UX evaluation approaches, and the growing recognition of UX as a multidimensional construct beyond usability. Prior studies predate these developments and offer only a limited representation of the current state of UX evaluation in government applications. With the focus on the 2023-2025 recent studies, this review intends to identify emerging UX evaluation approaches, UX dimensions, and contextual and cultural considerations that are relevant to contemporary government applications. This review contributes to the literature by analyzing current UX evaluation practices in government application studies and by highlighting the role of domain-specificity, contextual, and cultural considerations within these evaluations.

This study presents a systematic review of the literature on UX evaluation in government applications. The objectives of the review are to: 1) identify UX evaluation approaches and UX dimensions that are used in the UX evaluation of government applications, and 2) to examine how domains, and contextual and cultural considerations shape the UX evaluation of government applications. In particular, this review aims to contribute to the emerging body of scholarship on UX evaluation and digital government by providing a structured overview and synthesis of UX evaluation for government applications.

The remainder of this study is structured as follows: Section II details the review method, Section III presents the findings of the review, Section IV offers the discussion, and Section V concludes the study. The following section describes the review method undertaken in this systematic review.

II. Review Method

This systematic review was planned and executed according to the guidelines established by [12], which provide a well-defined structure for conducting systematic reviews within the computing context. For reporting the systematic review's findings, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram [13], [14] was employed. The combination of Kitchenham and Charters' guidelines [12] with PRISMA provides a methodologically rigorous, structured set of steps for defining review objectives, formulating Research Questions (RQs), and outlining

procedures for searching databases, selecting studies, extracting data, and synthesizing systematically.

As established by [12], the systematic review process has three phases: planning the review, conducting the review, and documenting/reporting the review. This study aligns with these three phases. There are specific activities for each phase, which are documented in Table I and further explained in the following subsections.

TABLE I. MAIN ACTIVITIES IN SYSTEMATIC REVIEW PHASES

Phase	Activities
Planning the review	Developing a review protocol
	Identifying research questions
	Identifying inclusion and exclusion criteria
Conducting the review	Implement search strategy and study selection
	Select primary studies
	Perform quality assessment
Documenting/reporting the review	Data extraction and synthesis

A. Planning the Review

This phase involves developing the review protocol, which includes identifying the RQs for this systematic review and establishing the inclusion and exclusion criteria.

1) Review protocol: Developing the review protocol is a fundamental first step in all systematic review protocols, as it provides a predefined plan that serves as a blueprint and a control measure to reduce bias introduced by the researchers [15], [12]. Following the Kitchenham and Charters guidelines, as demonstrated by [16], this study's review protocol is presented in Fig. 1:

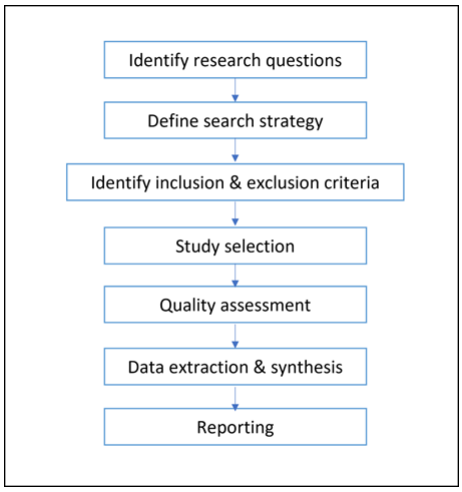

Fig. 1. Review protocol.

2) Research questions: This systematic review is anchored by two RQs that establish the foundation for this study:

RQ1: What UX evaluation approaches and UX dimensions have been employed in the UX evaluation of government applications?

RQ2: How do domains and contextual and cultural considerations influence the UX evaluation of government applications?

3) Search strategy: To define the search strategy, primary keywords and relevant terms are identified, and search strings are constructed using lexical resources (e.g., dictionaries and thesauri), reference materials, and prior scholarly works. Consequently, the search strings are entered into two major scholarly databases, Scopus and Web of Science (WoS), to search for relevant studies for the review process. Scopus and WoS are selected as the primary databases due to their robustness, high-quality indexing, comprehensive coverage, peer-reviewed status, and wide acceptance among the scholarly community. The search strings used are detailed in Table II. From this initial search, a total of 1,622 publications related to this study were identified, laying a solid foundation for further screening and analysis.

TABLE II. SEARCH STRINGS

Database	Search strings
Scopus	TITLE-ABS-KEY (("user experience evaluation" OR ux evaluation) AND (government OR "public service")) AND (LIMIT-TO (SUBJAREA , "COMP") OR LIMIT-TO (SUBJAREA , "ENGI") OR LIMIT-TO (SUBJAREA , "MATH") OR LIMIT-TO (SUBJAREA , "MULT") OR LIMIT-TO (SUBJAREA , "ECON") OR LIMIT-TO (SUBJAREA , "BUSI") OR LIMIT-TO (SUBJAREA , "DECI")) AND (LIMIT-TO (PUBYEAR , 2023) OR LIMIT-TO (PUBYEAR , 2024) OR LIMIT-TO (PUBYEAR , 2025)) AND (LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (LANGUAGE , "English")) Date of Access: May 2025
Web of Science	("user experience evaluation" OR ux evaluation) AND (government OR "public service") (Topic) and 2025 or 2023 or 2024 (Publication Years) and Article (Document Types) and English (Languages) and Article (Document Types) and English (Languages) Date of Access: May 2025

4) Inclusion and exclusion criteria: The next step in the review protocol is screening, where the initial set of 1,622 publications from the first stage is refined using inclusion and exclusion criteria. This refinement is conducted to ensure that the publications are related to this study and support the research objectives. Note that studies were excluded if they were published before 2023, written in a language other than English, or classified as conference papers, book chapters, review articles, or articles still in press. In addition, records belonging to subject areas outside the scope of the research, specifically computer science, engineering, business, management and accounting, economics, econometrics and finance, and multidisciplinary fields, were also removed. Following this rigorous screening, a total of 1,441 records were excluded. The remaining dataset comprised 149 publications, with 99 sourced from Scopus and 50 from WoS. A subsequent duplicate check was performed using a reference manager,

resulting in the removal of 18 duplicate records, leaving 131 publications for the next step of the review protocol. In essence, this process ensured that the final set of studies is both highly relevant and aligned with the research objectives, providing a robust foundation for the subsequent phases of the systematic review. Table III summarizes the inclusion and exclusion criteria used.

TABLE III. INCLUSION AND EXCLUSION CRITERIA FOR THE SELECTION CRITERIA IN SEARCHING

Criterion	Inclusion	Exclusion
Language	English	Non-English
Timeline	2023 – 2025	< 2023
Literature type	Journal (article)	Conference, Book, Review
Publication stage	Final	In press
Subject area	Computer science Engineering Business, management, and accounting Economics, econometrics, and finance Multidisciplinary	Besides Computer science Engineering Business, management, and accounting Economics, econometrics, and finance Multidisciplinary

5) Study selection: In this step, the titles and abstracts of the 131 publications from the previous phase are further examined to ensure eligibility, aligning with the established inclusion criteria and research objectives. A total of 97 articles were excluded on the grounds of falling outside the relevant field of UX evaluation in government applications, lacking significance based on their titles, presenting abstracts that did not correspond to the study's objectives, and/or lacking accessible full-text versions grounded in empirical research. Following this eligibility assessment, 34 articles were retained for quality assessment. Accordingly, the study process is documented using the PRISMA 2020 flow diagram [14], in Fig. 2.

6) Quality assessment: Quality assessment is an integral assessment in systematic reviews for determining the quality of the selected studies [12]. In this study, the quality assessment criteria are adopted from [17]'s quality assessment framework. There are a total of eight (8) quality assessment criteria formulated for this study. For the scoring system, a three-point score is assigned in which 1 ("Yes") is for when the criterion is fully satisfied, 0.5 ("Partially") is for when the criterion is only partially satisfied, and 0 ("No") is for when the criterion is not fully satisfied. The 34 articles from the previous step are screened using the quality assessment criteria. Following this screening, 15 articles were identified as not meeting 70% of the quality assessment criteria, leaving the remaining 19 articles as the primary studies for subsequent data extraction and analysis. Table IV presents the quality assessment criteria used in this step, and Table VI displays the results of the quality assessment of the selected papers.

TABLE IV. QUALITY ASSESSMENT CRITERIA

QA1	Is the purpose of the study clearly stated?
QA2	Is the interest and the usefulness of the work clearly presented?
QA3	Is the study methodology clearly established?
QA4	Are the concepts of the approach clearly defined?
QA5	Is the work compared and measured with other similar work?
QA6	Are the limitations of the work clearly mentioned?
QA7	Is the study explicitly focused on UX evaluation for government applications?
QA8	Does the study explicitly involve government applications or public service platforms as its domain of investigation?

7) Data extraction and synthesis: From the quality assessment step, nineteen studies are retained for data extraction and analysis. Data extraction is performed by examining each paper and extracting relevant data using MS Excel spreadsheets. This is done in accordance with guidelines [12] to construct data extraction forms to capture the data from the initial search in a systematic and reliable manner. Note that the selection of extracted data is guided by the RQs. It comprises the title of the article, authors, publication year, keywords, abstracts, journal titles, countries/regions in which the research is conducted, domains of investigation, platforms (web/mobile), UX evaluation approaches/methodologies, UX dimensions, and cultural and contextual considerations. The next section presents the findings of the review, organized according to descriptive overview and themes.

III. FINDINGS

The findings of this systematic review are outlined in this section, beginning with a descriptive overview of the studies, followed by a synthesis of the key themes identified across the reviewed studies. Discussion of the findings is presented in the next section.

A. Descriptive Overview

Publication years: The 19 papers in this study were published between 2023 and 2025, with the majority (10) published in 2024. Table V below summarizes the distribution of papers according to their year of publication.

TABLE V. PUBLICATION YEARS OF THE REVIEWED PAPERS

Year	Number of papers	Authors
2023	6	Yakovchenko et al. [18], Alharbi et al. [19], Hu et al. [20], Nagro [21], Cheng & Chen [22], Alghareeb et al. [23]
2024	10	Ro et al. [24], Benaida [25], Wahyuningrum et al. [26], Zubir & Latip [27], Puspasari et al. [28], Sapraz & Han [29], Nawafleh & Khasawneh [30], Aziz et al. [31], Fadrial et al. [32], Patergiannaki & Pollalis [33]
2025	3	Ashar et al. [34], Amer et al. [35], Kencono et al. [36]

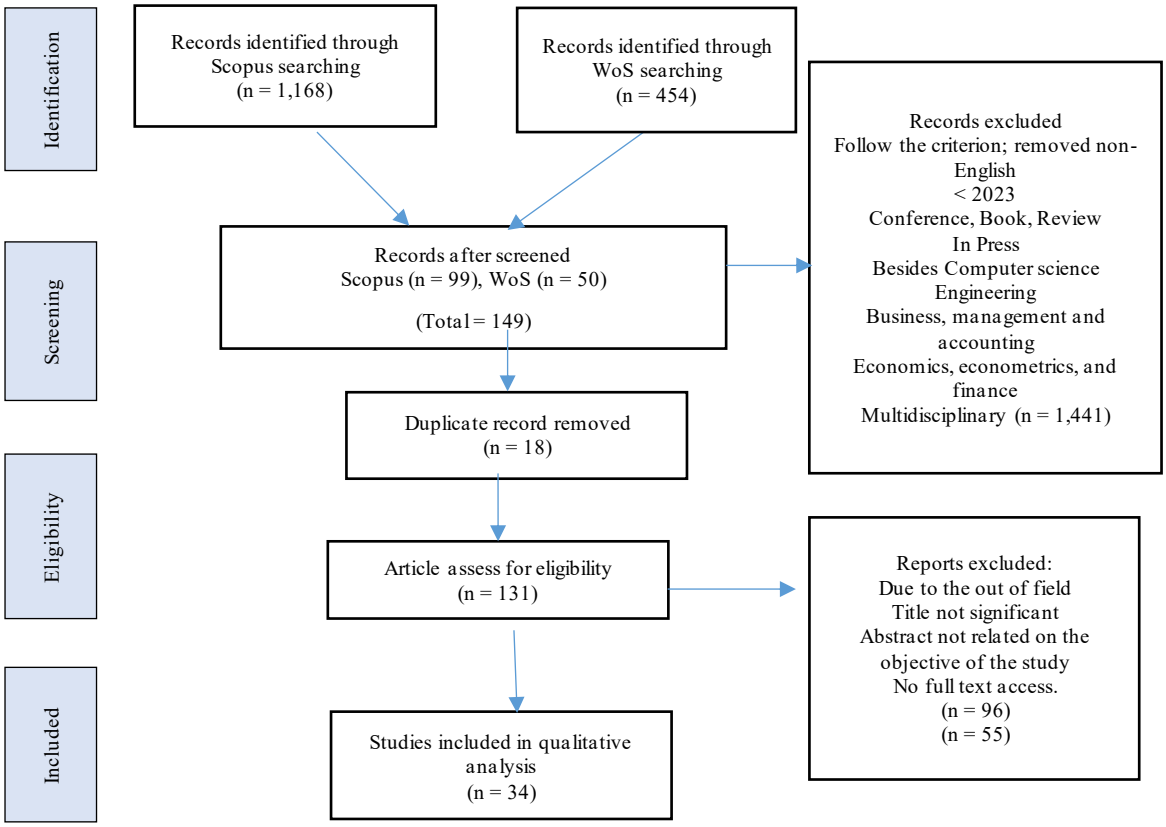

Fig. 2. The PRISMA flow diagram.

TABLE VI. QUALITY ASSESSMENT RESULTS OF THE SELECTED PAPERS

No.	Authors	Year	QA1	QA2	QA3	QA4	QA5	QA6	QA7	QA8	Total	%
1	Puspasari et al.	2024	1	1	1	0.5	0.5	0.5	1	1	6.5	81.3
2	Chi et al.	2024	1	1	1	1	1	0.5	0	0	5.5	68.8
3	Ro et al.	2024	1	1	1	1	0.5	1	1	1	7.5	93.8
4	Yakovchenko et al.	2023	1	1	1	1	1	1	1	1	8	100.0
5	Alharbi et al.	2023	1	1	1	1	1	1	1	1	8	100.0
6	Luo et al.	2024	1	1	1	0.5	0.5	0	0	0	4	50.0
7	Nagro S.A.	2023	1	1	1	1	1	0.5	1	1	7.5	93.8
8	Arya & Pal	2024	1	1	1	1	1	0.5	0	0	5.5	68.8
9	Cheng & Chen	2023	1	1	1	1	0.5	0	1	1	6.5	81.3
10	Geng et al.	2024	1	1	1	1	0.5	0.5	0	0	5	62.5
11	Křemen et al.	2025	1	1	0.5	1	0.5	0.5	0	0	4.5	56.3
12	Sapraz & Han	2024	1	1	1	1	1	0.5	1	1	7.5	93.8
13	Chang & Huang	2023	1	1	1	1	1	0.5	0	0	5.5	68.8
14	Sun et al.	2025	1	1	1	0.5	0.5	1	0	0	5	62.5
15	Nawafleh & Khasawneh	2024	1	1	1	1	0.5	0.5	1	1	7	87.5
16	Salama et al.	2023	1	1	1	1	1	0.5	0	0	5.5	68.8
17	Aziz et al.	2024	1	1	1	1	0.5	0.5	1	1	7	87.5
18	Koo et al.	2025	1	1	1	0.5	1	0.5	0	0	5	62.5
19	Benaida M.	2024	1	1	1	0.5	0.5	0.5			4.5	56.3
20	Norabuena-Mendoza et al.	2025	1	1	1	1	0.5	0.5	0	0	5	62.5

21	Wahyuningrum et al.	2024	1	1	1	1	0.5	0.5	1	1	7	87.5
22	Ashar et al.	2025	1	1	1	1	1	0.5	1	1	7.5	93.8
23	Mabkhot et al.	2024	1	1	1	1	1	0.5	0	0	5.5	68.8
24	Reddy et al.	2023	1	1	1	1	0.5	1	0	0	5.5	68.8
25	Amer et al.	2025	1	1	1	1	0.5	0.5	1	1	7	87.5
26	Fadrial et al.	2024	1	1	1	1	0.5	0.5	1	1	7	87.5
27	Alzebda et al.	2025	1	1	1	1	1	0.5	0	0	5.5	68.8
28	Kencono et al.	2025	1	1	0.5	0.5	0.5	0.5	1	1	6	75.0
29	Amro et al.	2023	1	1	1	1	0.5	0.5	0	0	5	62.5
30	Hu et al.	2023	1	1	1	1	0.5	0.5	1	1	7	87.5
31	Li et al.	2023	1	1	1	1	0.5	0.5	0	0	5	62.5
32	Alghareeb et al.	2023	1	1	1	1	1	1	1	1	8	100.0
33	Patergiannaki & Pollalis	2024	1	1	1	1	0.5	0.5	1	1	7	87.5
34	Zubir & Latip	2024	1	1	1	1	1	0.5	1	1	7.5	93.8

Domains covered: The 19 reviewed studies focused on government applications delivered via mobile and web platforms across six domains. The domains of the papers are listed in the following Table VII.

TABLE VII. DOMAINS COVERED IN THE REVIEWED PAPERS

Domain	Number of papers	Authors
Public services	9	Alharbi et al. [19], Benaida [25], Zubir & Latip [27], Nawafleh & Khasawneh [30], Fadrial et al. [32], Patergiannaki & Pollalis [33], Sapraz & Han [29], Amer et al. [35], Kencono et al. [36]
Healthcare	5	Yakovchenko et al. [18], Alghareeb et al. [23], Nagro [21], Wahyuningrum et al. [26], Ashar et al. [34]
Navigation	3	Hu et al.[20], Cheng & Chen [22]
Tourism	1	Aziz et al. [31]
Law enforcement	1	Puspasari et al. [28]
Assistive technology	1	Ro et al. [24]

Platforms: The studies are categorized into two platforms, mobile and web, as presented in Table VIII below.

TABLE VIII. PLATFORMS OF THE GOVERNMENT APPLICATIONS IN THE REVIEWED PAPERS

Platform	Number of Papers	Authors
Mobile	9	Hu et al. [20], Nagro [21], Alghareeb et al. [23], Wahyuningrum et al. [26], Zubir & Latip [27], Puspasari et al. [28], Aziz et al. [31], Ashar et al. [34], Kencono et al. [36]
Web	10	Yakovchenko et al. [18], Alharbi et al. [19], Cheng & Chen [22], Ro et al. [24], Benaida [25], Sapraz & Han [29], Nawafleh & Khasawneh [30], Fadrial et al. [32], Patergiannaki & Pollalis [33], Amer et al. [35]

Countries/regions: The studies were conducted in various regions, including Asia, Europe, Africa, and North America, with one cross-national study involving 50 countries [19]. Twelve countries in which the studies are conducted are presented in Table IX below. Two studies involved two countries in their works: Morocco and Tunisia [25] and Malaysia and Thailand [31].

TABLE IX. COUNTRIES/REGIONS OF THE GOVERNMENT APPLICATIONS IN THE REVIEWED PAPERS

Region	Number of countries	Country	Number of papers	Authors
Asia	9	Malaysia Indonesia Thailand South Korea China Sri Lanka Taiwan Saudi Arabia Jordan	3 4 1 1 1 1 1 2 2	Zubir & Latip [27], Aziz et al. [31], Ashar et al. [34] Wahyuningrum et al. [26], Puspasari et al. [28], Fadrial et al. [32], Kencono et al. [36] Aziz et al. [31] Ro et al. [24] Hu et al. [20] Sapraz & Han [29] Cheng & Chen [22] Nagro [21], Alghareeb et al. [23] Nawafleh & Khasawneh [30], Amer et al. [35]
Africa	2	Morocco, Tunisia	1	Benaida [25]
Europe	1	Greece	1	Patergiannaki & Pollalis [33]
North America	1	United States	1	Yakovchenko et al. [18]
Cross-national	1	50 countries	1	Alharbi et al. [19]

Evaluation approaches/methodologies: A mix of standardized usability/UX tools and methods, including quantitative, qualitative, and experimental/design-based nature, was used across the 19 studies. The following Table X illustrates the evaluation approaches/methodologies identified.

TABLE X. EVALUATION APPROACHES/METHODOLOGIES USED IN THE REVIEWED PAPERS

Evaluation approaches/methodologies	Authors
System Usability Scale (SUS) UMUX/UMUX-Lite User Experience Questionnaire (UEQ) Single Ease Question (SEQ) Research-based web design/usability guidelines (RBWDUG)	Yakovchenko et al. [18], Nagro [21] Wahyuningrum et al. [26], Nagro [21] Sapraz & Han [21] Wahyuningrum et al. [26] Benaida [25]
Expert reviews Interviews Think-aloud protocol Thematic analysis Sentiment analysis	Alharbi et al. [19], Aziz et al. [31] Cheng & Chen [22], Sapraz & Han [29], Amer et al. [35] Ro et al. [24] Ashar et al. [34], Puspasari et al. [28] Puspasari et al. [28], Kencono et al. [36]
Surveys and questionnaires Website analytics Process mining & social network analysis	Zubir & Latip [27], Nawafleh & Khasawneh [30], Fadrial et al. [32], Patergiannaki & Pollalis [33] Cheng & Chen [22], Ashar et al. [34] Amer et al. [35]
Value-Sensitive Design & FEDS framework AR usability experiments	Sapraz & Han [29] Hu et al. [20], Aziz et al. [31]

B. Themes

Examination of the reviewed studies revealed four central themes: diversity in UX evaluation approaches, the range of UX dimensions evaluated, the range of domains evaluated, and the consideration of contextual and cultural factors.

C. Diversity in UX Evaluation Approaches

The studies employed a diverse set of approaches to evaluate the UX of the web and mobile government applications. This demonstrates a methodological theme reflected in the use of qualitative, quantitative, and mixed-method approaches. For instance, [18], [21], [26], and [29] implemented standardized instruments for evaluating usability and UX, such as System Usability Scale (SUS), Usability Metric for User Experience (UMUX), UMUX-Lite, User Experience Questionnaire (UEQ), and Single Ease Questionnaire (SEQ). Meanwhile, [20] and [24] conducted their evaluations using task-based and experimental methods, including task analysis and behavior-emotion analysis. Inspection-based and heuristic evaluations were also incorporated, along with expert reviews and usability guidelines in [19], [31], and [25]. In addition, several studies combined mixed-methods approaches: analytics, surveys and interviews in [32], [34] and [35], sentiment and thematic analyses, and natural language processing of user reviews in [23], [28] and [36], and surveys, interviews, UEQ, and the Framework for Evaluation in Design Science Research (FEDS) in [29]. Moreover, [27], [30], [33], and [34] utilized large-scale surveys and are grounded in models such as the Technology Acceptance Model (TAM) and the Unified Theory of Acceptance and Use of Technology 2 (UTAUT2). At the same time, innovative approaches were also deployed, such as [35]'s use of a case study with process mining and network analysis. Essentially, the use of this diverse set highlights that the researchers rely on quantitative, qualitative, and mixed-methods approaches, enabling them to capture a range of UX dimensions during the evaluations.

D. Variety of UX Dimensions Evaluated

A variety of UX dimensions were evaluated in the reviewed studies. A recurring dimension is usability [19], [21], [18], [20], [25], [24], [26]. Effectiveness is a dimension investigated by [23], [24], [26], [28], [34], and [36]. Additionally, efficiency is also evaluated by [24], along with [23], [18], [25], and [35]. The satisfaction dimension is covered in the evaluations conducted by [23], [24], [25], [26], [30], and [36]. In [25], [27], and [33], the researchers investigated dimensions related to use: ease of use and perceived usefulness. Following this, accessibility is assessed by [19], [24], [25], and [26]. [19] also evaluated privacy, along with [36]. Conversely, [20] included emotional experience dimensions: pleasure, arousal, and dominance. Meanwhile, [32], [30], and [36] measured service quality, trust, security, and transparency. In [28] and [36], perceptions of performance, functionality, and performance were investigated. When considered together, the reviewed studies indicate that evaluations of UX encompass pragmatic, hedonic, and emotional responses, as well as government-specific concerns and context-specific cultural and social dimensions. Collectively, these highlight the multidimensional nature of UX in government applications.

E. Range of Domains in the UX Evaluations

The studies reviewed have undergone UX evaluation of government applications across various domains, including public services, healthcare, navigation, tourism, law enforcement, and assistive technology. This illustrates the varied functions provided by the government through its web and mobile platforms. The most prominent domain with the largest number of studies is public services delivered through government websites or mobile applications [19], [25], [27], [30], [32], [33], [29], [35], [36]. Several studies explicitly stated the type of public service provided. For example, [35]'s work evaluated electronic building permit systems, [29]'s work investigated a platform for citizens and the Sri Lankan government to collaborate on environmental issues, [33]'s work evaluated Greek municipal portals, and [36]'s work investigated the regional government mobile applications. The next domain is healthcare, which is investigated by [21], [18], [23], [26], and [34]. [20] and [22]'s works are in the navigation domain. Meanwhile, the domains of tourism, law enforcement, and assistive technology each have one study ([31], [28], [24]). The identified domains consist of established services, including public services, healthcare, tourism, and law enforcement, as well as emerging ones such as navigation and assistive technology. As a whole, these domains reflect the domain-specific needs that inform the UX evaluation criteria and shape the UX of the applications.

F. Contextual and Cultural Considerations in UX Evaluations

The studies were conducted across a diverse set of countries and regions, bringing the emergence of contextual and cultural considerations as a notable theme and reflecting their roles in UX evaluations. The majority of the works were situated in Asia: Malaysia [27], [34], [31], Indonesia [26], [28], [32], [36], China [20], Taiwan ([22]), South Korea [24], Sri Lanka [29], Saudi Arabia [21]; [23], and Jordan [30], [35]. Correspondingly, the Asian works highlighted issues such as accessibility, cultural

heritage integration, citizen trust, digital adoption, loyalty, usability, sentiment towards government applications, and the influence of social and cultural norms on the adoption of government applications. Conversely, [25]'s work addresses the accessibility challenges of government applications in Northern African countries, specifically Morocco and Tunisia. [33]'s work, set in Europe, examined the Greek e-government service quality. In North America, [18] investigated the large-scale United States health management dashboards' usability. Another work is cross-national, involving 50 countries' e-government websites' privacy norms and usability [19]. Collectively, these studies have proven that UX evaluation is shaped by both technical considerations and contextual and cultural considerations, as well as contexts from the institutional and socio-political landscapes. The UX evaluation of government applications is thus demonstrated to involve contextual considerations, local cultural expectations, policy frameworks, and societal norms. This ultimately affirms the importance of contextually and culturally grounded evaluation approaches in conducting UX evaluations. These findings are discussed in the subsequent section.

IV. Discussion

This systematic review sets out to address two RQs:

RQ1: What UX evaluation approaches and UX dimensions have been employed in the UX evaluation of government applications?

RQ2: How do domains and contextual and cultural considerations influence the UX evaluation of government applications?

On the first RQ, a core finding that is consistently observed throughout the studies is the diverse approaches used in UX evaluations of government applications. Furthermore, a wide range of approaches is applied by the researchers. It ranges from standardized instruments for usability and UX evaluation (e.g., SUS, UMUX, UEQ), analytical and interpretive techniques (task analysis, behavior-emotion analysis, heuristic evaluations, expert reviews, surveys, and interviews), computational and data-driven methods (sentiment analysis, thematic analysis, natural language processing of user reviews, process mining, and network analysis), to framework-based approaches such as the FEDS ([18], [21], [26], [29], [20], [24], [19], [31], [25], [32], [34], [35], [23], [28], [36], [27], [30], [33], [35]. In addition, the diversity in the UX evaluation approaches deployed by the researchers is two-pronged. However, this current practice is both rich and fragmented in the sense that while various approaches can assess a variety of UX dimensions, the lack of a widely-recognized or dominant single approach suggests that standardization in UX evaluation for government applications has yet to be established. Similarly, the variety of UX dimensions evaluated across the studies has also revealed a significant pattern. Pragmatic dimensions such as usability, efficiency, effectiveness and satisfaction are consistently examined [19], [21], [18], [20], [23], [25], [24], [26], [28], [30], [27], [33], [34], [36], [35] but hedonic and emotional dimensions [20] are limited. Meanwhile, government-specific dimensions such as privacy, transparency, trust, and service quality appeared as notable dimensions in several studies [25], [30], [36]. This signifies a progression from evaluations focused purely on usability, oriented to a richer understanding of UX. Nevertheless, gaps emerged from the uneven distribution of attention across the dimensions. Cultural, emotional, and social dimensions remain underexplored compared to the pragmatic or functional dimensions.

Moving to RQ2, the reviewed studies indicate that domains, as well as contextual and cultural considerations, influence UX evaluations in government applications. More nuance is gained through the discussion of the range of government applications in the reviewed studies. This range reflects the expanding role of government applications in mediating interaction between governments and their citizens. The domains identified comprise conventional services such as public services, healthcare, tourism, and law enforcement [19], [25], [27], [30], [32], [33], [29], [35], [36], [18], [23], [21], [26], [34], [31], [28], together with emerging areas such as navigation and assistive technology [20], [22], [24]. In line with this, it is evident that domain-specific needs inform evaluation, e.g., public services prioritize service quality [32], [30], [36], [33], [25], [30], and healthcare applications prioritize data security and trust [21], [18]. This highlights the importance of considering the domain of the government application being evaluated. Another defining theme across the reviewed studies is contextual and cultural considerations in UX evaluations. Studies situated in Asian settings underscore the significance of accessibility, loyalty, and trust, signaling that culture and social norms influence adoption ([26], [29], [36]. Outside Asia, [18] in North America, [25]'s study, and [33]'s study in Europe drew attention to accessibility and service quality. These variations suggest that UX in government applications is not mutually exclusive from the interplay of socio-political and cultural settings. Despite this, few studies integrate cultural dimensions into their UX evaluations, marking a potential direction for further investigation into the field.

The progress of standardized yet adaptable frameworks for UX evaluation in government applications should be placed at the forefront of scholarly investigation. While best practices can be consolidated, they must remain flexible in varied cultural and domain environments. Additionally, there is a strong need to incorporate hedonic and emotional dimensions, considering their influence on the trust and loyalty of the citizen users. This supplements the common pragmatic dimensions, such as usability. Hence, contextual and cultural considerations should be more clearly integrated into UX evaluations, which could improve these UX evaluations across various settings. Domain-specific strategies should also be considered in future work, as UX evaluations are informed and shaped by domain-specific needs. Together, these directions will contribute to more comprehensive, contextually grounded, and culturally sensitive UX evaluation for government applications.

Collectively, the findings of this review give rise to several meaningful implications for UX evaluation in government applications. For the research community, the findings underscore the necessity to adopt more comprehensive perspectives that account for UX's multifaceted nature and to move away from usability-centric evaluations in conducting UX evaluations in government applications. For designers, developers and UX practitioners involved in building government applications, the findings suggest that UX

evaluation criteria should be adapted to include the domain, service and user contexts, without using only generic usability or UX evaluations. For governments, the findings show that UX evaluation with domain, contextual and cultural considerations can be accommodated earlier in the government application development lifecycle, to help realize more effective, inclusive, trustworthy and citizen-centric applications. The subsequent section provides a conclusion of this study.

V. CONCLUSION

Four key themes emerged from this systematic review of UX evaluations for government applications, guided by two RQs: diversity in UX evaluation approaches, the variety of UX dimensions evaluated, the range of domains in UX evaluations, and the contextual and cultural considerations in UX evaluations. These themes indicate the richness and fragmentation of current practices. Pragmatic dimensions, such as usability, remain the main focus of UX evaluations, while hedonic, emotional, and cultural dimensions are underexplored. Concurrently, despite the diverse approaches employed, there is a lack of standardized yet adaptable frameworks for UX evaluations in government applications. Furthermore, this systematic review highlights that UX evaluations in government applications should encompass both technical and pragmatic dimensions, as well as domain-specific, cultural, and contextual aspects. The implications of the findings for research, practice and stakeholder communities outline how UX evaluation in government applications may be strengthened. Therefore, strengthening UX evaluations through this approach will ultimately enable more inclusive and meaningful evaluations, likely resulting in government applications with a positive UX.

ACKNOWLEDGMENT

The authors would like to thank Universiti Malaysia Sarawak, the Ministry of Higher Education Malaysia, and Universiti Putra Malaysia for the valuable support in this study.

REFERENCES

[1] D. Norman and J. Nielsen, "The Definition of User Experience (UX)." Accessed: May 01, 2025. [Online]. Available: https://www.nngroup.com/articles/definition-user-experience/

[2] N. Yusof, N. L. Hashim, and A. Hussain, "A Conceptual User Experience Evaluation Model on Online Systems," Int. J. Adv. Comput. Sci. Appl., no. May, 2022, doi: 10.14569/IJACSA.2022.0130153.

[3] N. L. Hashim, N. Yusof, P. B. Pulau, A. Hussain, and M. Ibrahim, "User Experience Dimensions for E-procurement : A Systematic Review," J. Inf. Commun. Technol., no. January 2023, 2022, doi: 10.32890/jict2022.21.4.1.

[4] H. Sharp, Y. Rogers, and J. Preece, Interaction Design: Beyond Human-Computer Interaction, 5th Editio. Wiley.

[5] Y. Rogers, H. Sharp, and J. Preece, Interaction Design: Beyond Human-Computer Interaction, 6th Editio. Wiley, 2023.

[6] J. Lewis and J. Sauro, "Three Branches of Standardized UX Measurement." Accessed: May 01, 2025. [Online]. Available: https://measuringu.com/three-branches-ux/

[7] R. M. Ramli, "E-GOVERNMENT IMPLEMENTATION CHALLENGES IN MALAYSIA AND SOUTH KOREA: A COMPARATIVE STUDY," Electron. J. Inf. Syst. Dev. Ctries., vol. 80, no. 7, pp. 1–26, 2017.

[8] N. S. Abdul Wahi, "The current development of e-government in Malaysia," in A HAZA SZOLGÁLATÁBAN KONFERENCIAKÖTET 2021., A. Kiss, Rebeka; Urbanovics, Ed., 2021.

[9] A. Aldrees and D. Gracanin, "UX in E-government Services for Citizens : A Systematic Literature Review," J. User Exp., vol. 18, no. 3, pp. 133–169, 2023.

[10] I. Lyzara, Ria; Purwandari, Betty; Zulfikar, Muhammad Fadhil; Santoso, Harry Budi; Solichah, "E-Government Usability Evaluation: Insights from A Systematic Literature Review," in Proceedings of the 2019 2nd International Conference on Software Engineering and Information Management, ICSIM 2019, Bali, Indonesia: ACM International Conference Proceeding Series, 2019, pp. 249–253. doi: https://doi.org/10.1145/3305160.3305178.

[11] N. P. Bastardo, Rute; Pavão, João; Rocha, "Methodological Quality of User-Centered Usability Evaluation of Digital Applications to Promote Citizens ' Engagement and Participation in Public Governance : A Systematic Literature Review," pp. 740–761, 2024.

[12] B. Kitchenham and S. Charters, "Guidelines for performing Systematic Literature Reviews in Software Engineering," 2007.

[13] D. G. Moher, David; Liberati, Alessandro; Tetzlaff, Jennifer; Altman, "Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement," BMJ 2009;339b2535, 2009, doi: https://doi.org/10.1136/bmj.b2535.

[14] M. J. Paige et al., "The PRISMA 2020 statement : an updated guideline for reporting systematic reviews Systematic reviews and Meta-Analyses," BMJ 2021;372n71, 2021, doi: 10.1136/bmj.n71.

[15] B. Kitchenham, "Procedures for Performing Systematic Reviews," 2004.

[16] S. Asadi, R. Abdullah, Y. Yah, and S. Nazir, "Understanding Institutional Repository in Higher Learning Institutions : A Systematic Literature Review and Directions for Future Research," IEEE Access, vol. 7, pp. 35242–35263, 2019, doi: 10.1109/ACCESS.2019.2897729.

[17] A. Abouzahra, A. Sabraoui, and K. Afdel, "Model composition in Model Driven Engineering : A systematic literature review," Inf. Softw. Technol., vol. 125, no. April, p. 106316, 2020, doi: 10.1016/j.infsof.2020.106316.

[18] V. Yakovchenko, D. A. Jacob, S. S. Rogal, T. R. Morgan, and K. Rozenberg-Ben-Dror, "User experience of a hepatitis c population management dashboard in the Department of Veterans Affairs," PLoS One, vol. 18, no. 5 May, 2023, doi: 10.1371/journal.pone.0285044.

[19] J. A. Alharbi, A. S. Albesher, and H. A. Wahsheh, "An Empirical Analysis of E-Governments' Cookie Interfaces in 50 Countries," Sustain., vol. 15, no. 2, 2023, doi: 10.3390/su15021231.

[20] S. Hu, L. D. Rong, J. L. Han, D. Zhang, and W. T. Jiang, "The Effects of Interaction Mode and Individual Differences on Usability and User Experience of Mobile Augmented Reality Navigation," IEEE ACCESS, vol. 11, pp. 41783–41795, 2023, doi: 10.1109/ACCESS.2023.3271522.

[21] S. A. Nagro, "User Experiences and Usability Evaluation of COVID-19 Application," J. Comput. Sci., vol. 19, no. 3, pp. 372–388, 2023, doi: 10.3844/JCSSP.2023.372.388.

[22] Y.-J. Cheng and K.-H. Chen, "Website analytics for government user behavior during COVID-19 pandemic," Aslib J. Inf. Manag., vol. 75, no. 1, pp. 90–111, 2023, doi: 10.1108/AJIM-11-2021-0329.

[23] M. Alghareeb, A. S. Albesher, and A. Asif, "Studying Users' Perceptions of COVID-19 Mobile Applications in Saudi Arabia," SUSTAINABILITY, vol. 15, no. 2, 2023, doi: 10.3390/su15020956.

[24] E. R. Ro, K. O. An, A. J. Kim, S. U. Jang, E. J. Kim, and S. D. Eun, "Usability Study to Promote Co-Creation among People with Disabilities, Developers, and Makers with a Focus on the Assistive Technology Open Platform in Korea," IEEE Access, vol. 12, pp. 39016–39027, 2024, doi: 10.1109/ACCESS.2023.3345036.

[25] M. Benaida, "Usability Evaluation of North African E-Government Services in the Context of Optimizing User Experience when Using Mobile Phones," Int. J. Interact. Mob. Technol., vol. 18, no. 8, pp. 102–115, 2024, doi: 10.3991/ijim.v18i08.46329.

[26] T. Wahyuningrum et al., "Logistic Regression Analysis of Factors that Influence User Experience in Student Medical Report Applications," J. Appl. Data Sci., vol. 5, no. 3, pp. 1210–1222, 2024, doi: 10.47738/jads.v5i3.285.

[27] M. H. H. Zubir and M. S. A. Latip, "Factors affecting citizens' intention to use e-government services: assessing the mediating effect of perceived usefulness and ease of use," Transform. Gov. PEOPLE Process POLICY, vol. 18, no. 3, pp. 384–399, 2024, doi: 10.1108/TG-04-2023-0040.

[28] H. M. Puspasari, I. Z. Mustaqim, A. T. Utami, R. Syalevi, and Y. Ruldeviyani, "Evaluation of Indonesia's police public service platforms through sentiment and thematic analysis," IAES Int. J. Artif. Intell., vol. 13, no. 2, pp. 1596–1607, 2024, doi: 10.11591/ijai.v13.i2.pp1596-1607.

[29] M. Sapraz and S. Han, "Users' evaluation of a digital government collaborative platform (DGCP) in Sri Lanka," Transform. Gov. People, Process Policy, vol. 18, no. 1, pp. 131–144, 2024, doi: 10.1108/TG-10-2022-0140.

[30] S. Nawafleh and A. Khasawneh, "Drivers of citizens E- loyalty in E-government services: E-service quality mediated by E-trust based on moderation role by system anxiety," Transform. Gov. People, Process Policy, vol. 18, no. 2, pp. 217–240, 2024, doi: 10.1108/TG-04-2023-0053.

[31] N. Aziz, S. S. Hamzah, S. Z. Ahmad, W. Matcha, and S. Binsaleh, "Augmented Reality Mobile Application for Promoting Culture and Heritage in Thailand and Malaysia: The Prototype Development and Heuristic Evaluation," Int. J. Interact. Mob. Technol. , vol. 18, no. 14, pp. 72–89, 2024, doi: 10.3991/ijim.v18i14.48381.

[32] R. Fadrial, H. T. R. Freddy Simanjuntak, W. Wirman, and W. S. Wibowo, "FOSTERING TRUST THROUGH BYTES: UNRAVELLING THE IMPACT OF E-GOVERNMENT ON PUBLIC TRUST IN INDONESIAN LOCAL GOVERNMENT," Interdiscip. J. Information, Knowledge, Manag., vol. 19, 2024, doi: 10.28945/5317.

[33] Z. Patergiannaki and Y. A. Pollalis, "E-government quality from the citizen's perspective: the role of perceived factors, demographic variables and the digital divide," Int. J. PUBLIC Sect. Manag., vol. 37, no. 2, pp. 232–254, 2024, doi: 10.1108/IJPSM-07-2023-0229.

[34] A. M. Ashar, N. F. Elias, R. Jenal, M. C. Lam, M. R. Appannan, and S. A. Iskandar, "Evaluating and Optimizing MySejahtera App Analytics for Sustainable Digital Government Services," IEEE Access, 2025, doi: 10.1109/ACCESS.2025.3561357.

[35] L. Amer, M. Thneibat, F. Sammour, and N. Lepkova, "EVALUATING THE E-PERMIT SYSTEM IN CONSTRUCTION USING STAKEHOLDER ANALYSIS AND NETWORK THEORY," J. Civ. Eng. Manag., vol. 31, no. 3, pp. 224–247, 2025, doi: 10.3846/jcem.2025.22253.

[36] D. S. Kencono, A. Djunaedi, and Y. Purbokusumo, "Exploring Public Sentiment Toward Regional E-Government Apps: A Case Study of Sapawarga and Jaki in Indonesia," J. Inf. Syst. Eng. Manag., vol. 10, pp. 285–298, 2025, doi: 10.52783/jisem.v10i15s.2458.

[37] B. Richardson, M. Campbell-yeo, and M. Smit, "Mobile Application User Experience Checklist : A Tool to Assess Attention to Core UX Principles Mobile Application User Experience Checklist : A Tool to Assess Attention to Core UX," Int. J. Human–Computer Interact., vol. 00, no. 00, pp. 1–8, 2021, doi: 10.1080/10447318.2021.1876361.

[38] Q. Zheng and M. Chen, "EvAlignUX : Advancing UX Evaluation through LLM-Supported Metrics Exploration," 2025, doi: 10.1145/3706598.3714045.

[39] S. Liu, H. Zhang, J. U. N. Kong, L. Zhang, and C. Gao, "UXBIV : An Evaluation Framework for Business Intelligence Visualization," IEEE Access, vol. 11, no. August, pp. 92391–92415, 2023, doi: 10.1109/ACCESS.2023.3300418.

[40] H. Jang and S. H. Han, "International Journal of Human - Computer Studies User experience framework for understanding user experience in blockchain services," Int. J. Hum. - Comput. Stud., vol. 158, no. October 2021, p. 102733, 2022, doi: 10.1016/j.ijhcs.2021.102733.

[41] A. P. Gallegos, G. R. Morales, P. V Torres-Carrión, and S. C. Carrión, "Framework for User Experience Evaluation in MOOC Platforms," in HCII 2022, LNAI 13336, H. Degen and S. Ntoa, Eds., Springer, 2022, pp. 284–304.

[42] N. F. Syahrozad and A. P. Subriadi, "Evaluation of user experience : a systematic literature review," vol. 02009, pp. 1–7, 2024.

[43] A. I. Nur, H. B. Santoso, and P. O. Hadi Putra, "The Method and Metric of User Experience Evaluation : A Systematic Literature Review," pp. 307–317, 2017, doi: 10.1145/3457784.3457832.

[44] V. G. de Menezes, G. V Pedrosa, M. P. P. Silva, and R. M. C. Figueiredo, "Evaluation of Public Services Considering the Expectations of Users — A Systematic Literature Review," pp. 1–13, 2022.