A Domain-Independent Framework for Effective Prioritization and Evaluation of UX Aspects in Mobile Apps

Haifa Alshammare
Information and Computer Science Department, KFUPM, Dhahran, Saudi Arabia, *Digital Technical College for Girls, Technical and Vocational Training Corporation*, Al-Ahsa, Saudi Arabia,
g202110890@email.com

Mohammad Alshayeb
Information and Computer Science Department, KFUPM, Dhahran 31261, Saudi Arabia, *Interdisciplinary Research Center for Intelligent Secure Systems*, Dhahran 31261, Saudi Arabia,
alshayeb@kfupm.edu.sa

Malak Baslyman
Information and Computer Science Department, KFUPM, Dhahran 31261, Saudi Arabia, *Interdisciplinary Research Center for Finance and Digital Economy,* Dhahran 31261, Saudi Arabia,
Malak.baslyman@kfupm.edu.sa

***Abstract*—The rapid evolution of mobile applications has intensified the need to prioritize and systematically evaluate user experience (UX) aspects. This paper introduces a domain-independent hybrid framework that ranks UX aspects and their evaluation methods. The framework integrates the MOSCOW method and the Analytic Hierarchy Process (AHP) to evaluate UX based on User Profiles, App Domain Specificity, Task Complexity, and Feasibility. A Use Value–Effort/Complexity matrix further supports the evaluation of methods by considering evaluation goals, user profiles, application domain, stage of development, data needs, resource availability, technical complexity, and ease of data collection and analysis. Six evaluators validated the framework with a case study, including three domains: mobile banking, gaming, and health. The framework enabled them to agree on core UX priorities, demonstrating its flexibility and practicality, offering a clear method for prioritizing and evaluating UX across mobile app domains.**

***Keywords—Interactive systems, Mobile applications, User experience, HCI, UX aspects, UX Evaluation methods, MOSCOW, Analytic Hierarchy Process (AHP), Use Value-Effort/Complexity*.**

I. INTRODUCTION

Mobile apps serve users across diverse domains and are increasingly expected to deliver high-quality UX [1]. As a multifaceted concept, UX plays a central role in shaping user satisfaction, retention, and overall app success [2] [3][4] [5] [6] [7]. It is influenced by the user's internal state, system characteristics, and interaction context [8]. Different types of mobile apps emphasize different UX priorities. For example, shopping apps may emphasize visual aesthetics, while business apps often prioritize efficiency [9]. However, systematic studies of how UX priorities vary by domain remain limited [9] [10]. highlighting the need for structured frameworks to prioritize UX aspects and their evaluation methods. Much like requirement prioritization in software engineering, selecting a few key UX aspects is a straightforward process. However, managing a wider range of UX aspects and choosing suitable evaluation methods becomes complex [11]. Prioritizing helps teams make better decisions, use resources efficiently, and focus on UX evaluation for mobile app development. Not all UX aspects or evaluation methods are equally relevant or feasible across different mobile app domains [11]. Several prioritization techniques have been proposed in software engineering, including AHP, MOSCOW, Cumulative Voting (CV), and fuzzy logic, each with its own advantages and limitations [11]. For example, AHP offers consistency and structure through pairwise comparisons, but it becomes unwieldy with large item sets [11]. The MOSCOW method offers a straightforward approach by categorizing items as "Must," "Should," "Could," or "Won't," yet it lacks the precision of more mathematically grounded methods [11]. Therefore, hybrid approaches are better suited to strike a balance between usability and rigor. Hence, this paper proposes a domain-independent framework that combines the MOSCOW method with AHP to address these challenges. First, MOSCOW reduces the number of UX aspects, simplifying the prioritization process. Then, AHP ranks the remaining aspects through pairwise comparisons, providing a structured and systematic decision-making method. AHP is particularly well-suited for multi-criteria decision-making [12] and has been extensively applied in human-computer interaction (HCI) research, especially for prioritizing usability criteria and assessing UX problems [12]. Additionally, we introduce the Use Value-Effort/Complexity Matrix to prioritize the evaluation methods of the UX aspects. This framework assesses the priority of UX methods based on two axes: value and implementation effort [12] [9]. By plotting each item on this matrix, development teams can make informed, strategic decisions about which initiatives to prioritize based on their potential value and the complexity of their implementation [9] which gives the organizations a quantifiable reasoning framework [13]. The framework considers key criteria such as user profiles, application domain, task complexity, and feasibility to tailor decisions to specific contexts. Previous studies have shown that the type of product or service, user profiles, and the context of use are critical factors influencing UX [14] [15]. The study (framework) aims to: (1) Facilitate the identification of core UX aspects across various mobile app domains; (2) Provide a structured method to select suitable UX evaluation methods; (3) Reduce bias and ad-hoc decisions in UX planning and evaluation by considering key criteria. To validate the framework, we conducted a case study involving six evaluators across three mobile domains: banking, gaming, and health. The framework helped evaluators systematically prioritize UX aspects and select feasible, high-impact evaluation methods.

2151-0849/25/$31.00 ©2025 IEEE
DOI 10.1109/ASEW67777.2025.00029

The remainder of this paper is organized as follows: Section II reviews related work. Section III describes the proposed framework. Sections IV, V, and VI present the case study, discussion and implications, and threats to validity. Section VII concludes with directions for future work.

II. Related Work

Several studies have employed multi-criteria decision-making (MCDM) methods to prioritize UX. For example, Gutiérrez et al. [16] combined Fuzzy Cognitive Maps and DEMATEL to rank quality attributes in augmented reality apps, demonstrating the value of structured prioritization. Moreover, Klein et al. [9] utilized the Kano model to categorize and prioritize UX elements of voice user interfaces (VUIs) obtained from user data through a survey, while Deutschländer et al. [17] concentrated on assessing the UX elements of VUIs, mainly focusing on the variation of these elements among different user groups, such as age. The researchers explored potential age-specific characteristics to determine the impact of age on UX prioritization. Employing the Kano model with age segmentation, they categorized 32 UX elements based on data collected from 384 VUI users. Similarly, Malinka et al. [18] utilized the Kano model to prioritize quality principles for health apps, gathering input from physicians. Additionally, user surveys are among the methods used for prioritization. Schrepp et al. [18] surveyed users across product categories, identifying the most critical UX elements for design.Likely, Klingbeil et al. [19] explored the prioritization of UX elements in learning environments to measure user satisfaction by integrating them into a questionnaire. Concerning the AHP technique, Dhouib et al. [12] addressed the prioritization of usability criteria, utilizing the AHP in adaptive user interfaces of information systems, thereby providing a structured framework for evaluating adaptive systems. While studies have not yet prioritized UX evaluation methods, Rajeshkumar and Omar [20] proposed a taxonomy of 89 UX evaluation methods (UXEMs), organized by study type, phase, and evaluator role. Although this research did not center on prioritization, it provided a framework for selecting appropriate evaluation methods for different development stages, supporting the use of relevant methods in interaction design. However, most prior studies are domain-specific (e.g., learning [19], health [18], VUIs [17] [9]) or focus only on usability [12]. Furthermore, in the augmented reality (AR) field, studies such as[16] concentrate on weighting system attributes rather than comparing different alternatives. Additionally, some studies rely on questionnaires as the primary tool for prioritization [21], and one study integrates methods like AHP [11] within the MOSCOW technique for requirements engineering prioritization, without including any specific criteria. None, however, systematically prioritizes UX evaluation methods. Consequently, a domain-independent, systematic approach to prioritizing and ranking UX aspects and evaluation methods across various context criteria is needed to advance research and practice in the field.

III. Proposed framework

This paper presents a domain-independent hybrid framework comprising two main phases that could be used separately (as illustrated in Fig. 1): the first focuses on ranking UX aspects for evaluation from a comprehensive list, and the second prioritizes the evaluation methods available for these aspects. The framework has been automated using Excel[1] to facilitate this process, making it accessible and easy to apply.

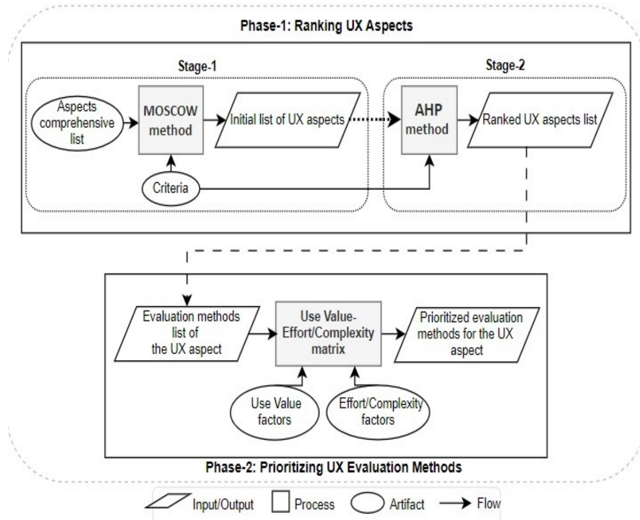

Fig. 1. Proposed framework.

A. *Ranking UX Aspects*

Our framework combines the MOSCOW method [11] and the AHP [22] to systematically rank UX aspects. The process begins with Stage-1, where each UX aspect is categorized using the MOSCOW method into one of four priority groups: (1)**Must-have** aspects are of the highest priority and are critical for the app's functionality and user satisfaction; (2) **Should-have** aspects are important for enhancing UX but are not essential for the app's core functionality; (3) **Could-have** aspects are non-essential but can further improve UX; (4) **Won't-have** aspects hold the lowest priority due to irrelevance, low impact, or existing constraints.

Each UX aspect is assessed across four criteria—User Profiles, App Domain Specificity, Task Complexity, and Feasibility (Table I)—recognized as key in human-technology interaction [15] [23] [24]. Aspects rated as "Must-have" or "Should-have" in at least one criterion are retained. This filtering step reduces complexity [11] by excluding non-essential aspects and focusing on those most relevant for the next stage.

Table I. Criteria used in Stage-1 (MOSCOW) and Stage-2 (AHP).

Criteria	Definition
User profiles	The importance of this aspect for the different user groups of the mobile app.
App domain specificity	The relevance of this aspect to the mobile apps domain will be developed or evaluated.
Task complexity	The importance of this aspect for assessing complex tasks.
Feasibility	The feasibility of implementing/evaluating this aspect, given available resources (time, cost, personnel).

[1] Aspects and evaluation methods prioritization ranking tool (https://github.com/PapersRepository/Framework-1-2-/blob/117f5535dc82a0d923f05e2ec6981f9c17616df9/Aspects%20ranking%20tool.xlsx)

In Stage-2, the initial list of UX aspects is ranked using the AHP method, as outlined in [22]. The AHP hierarchy is structured into two levels: the first level comprises the criteria used in the MOSCOW method (user profiles, app domain specificity, task complexity, and feasibility), and the second level includes the alternatives, which are the UX aspects derived from Stage-1. During the criteria pairwise comparison step, evaluators assess the relative importance of each criterion using a nine-point scale, ranging from "equally important" (ratio = 1) to "extremely important" (ratio = 9). These comparisons automatically generate the criterion weights. Following this, the evaluator rates the alternatives (UX aspects) against each criterion using the same nine-point scale. The alternatives' scores are generated automatically based on these evaluations. As a result, the framework produces a ranked list of UX aspects by combining the criterion weights with the scores of the alternatives. Since the process is expert-led, evaluator differences are expected; the ranking makes them explicit and provides a defensible basis for selection. In practice, we select the smallest top-K list that covers all criteria and achieves a cumulative weight of ~60–70%, ensuring that domain-specific needs (e.g., Safety in regulated sectors) are met. This creates a tailored evaluation set while avoiding ad hoc choices. External factors such as UX maturity or compliance may also be considered. Future work will formalize the selection process and test various thresholding strategies to achieve consistent, context-sensitive prioritization.

B. Prioritizing UX Evaluation Methods

In the second part of our framework, we prioritize UX evaluation methods for evaluating the ranked UX aspects using a Use Value-Effort/Complexity matrix [13]. As illustrated in Fig. 2, the matrix is structured along two axes: the Use Value axis, which measures the method's significance for evaluating a specific UX aspect, and the Effort/Complexity axis, which captures the complexity or effort required to apply the method. The matrix is divided into four quadrants, each offering insights into method prioritization:

- Upper Left Quadrant: Methods in this quadrant exhibit high Use Value and low Effort/Complexity, making them highly efficient and ideal candidates for prioritization due to their high impact and minimal resource demands.
- Lower Right Quadrant: Methods in this quadrant have low Use Value and high Effort/Complexity, indicating that they are resource-intensive and offer limited benefit, suggesting they should be deprioritized or avoided unless necessary.
- Other Quadrants: Methods in these quadrants can be evaluated based on specific needs and constraints.

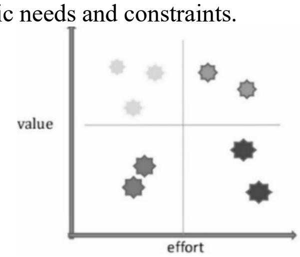

Fig. 2. Example of Use Value-Effort/Complexity Matrix [13].

Several factors influence the positioning of methods within the Use Value–Effort/Complexity matrix. For Use Value, five aspects guide the assessment. (1) **Evaluation goals** direct how well the method aligns with specific objectives, such as assessing experiences before use, capturing snapshots during interaction, evaluating tasks or activities, or examining long-term UX [20]; (2) **User profiles** indicate whether the method effectively accommodates the characteristics of the target users, such as novices versus experts; (3) **The application domain** determines the method's suitability for the type of app being assessed (e.g., e-commerce compared to health apps); (4) **The stage of development** [20] reveals how effectively the method serves the current phase, whether during prototyping or post-launch evaluation; (5) **The type of data** required determines whether the method yields qualitative insights (e.g., user perceptions) or quantitative metrics (e.g., performance measures). For Effort/Complexity, three main factors play a role: (1) **Resource availability** shapes constraints such as budget, time, and personnel [20]; (2) **Technical complexity** refers to the difficulty of implementing the method, including the tools, platforms, or setup required; (3)**The ease of data collection and analysis** determines how straightforward it is to gather and interpret data once the method is applied.

Each method is rated on these Use Value and Effort/Complexity factors using a 1–5 scale (1 = very low value/effort, 5 = very high). These ratings are plotted on the matrix. Priority is calculated by finding the ratio of Use Value to Effort; higher ratios mean higher priority. Methods with high Use Value and low Effort (e.g., a ratio of 5 for Use Value 5 and Effort 1) are prioritized. Methods with low Use Value and high Effort (e.g., a ratio of 0.25) are deprioritized. This system ensures that UX evaluations focus on impactful and feasible methods, thereby boosting assessment efficiency.

IV. CASE STUDY

To validate the effectiveness and flexibility of the proposed framework for prioritizing UX aspects in mobile apps, we conducted a case study involving six evaluators, each focused on a specific domain: banking, gaming, or health apps. For banking, Evaluators 1 and 2 hold degrees in software engineering and have experience in HCI. In the gaming domain, Evaluators 3 and 4, experienced in UX design and mobile development, focused on gaming apps. For health apps, Evaluators 5 and 6, both UX professionals with expertise in HCI and product design, concentrated on mHealth applications. The evaluators utilized the framework by applying the MOSCOW method and the AHP to prioritize relevant UX aspects for their respective domains. They utilized an automated Excel tool developed for Phase One, which encompasses Stages 1 and 2, as referred to in Section III. The tool integrates UX aspects from a comprehensive list derived from a systematic mapping study [25]. The primary goal of the case study was to evaluate the framework's effectiveness in identifying key UX aspects, ensuring its relevance and practical applicability for assessing and prioritizing UX aspects in real-world contexts.

A. Results of the UX Aspects Ranking Approach

1) MOSCOW Method Results

As shown in TABLE *II*, Evaluators-1 and Evaluator-2, who assessed mobile banking apps, identified key UX aspects such as Usability, Safety, Accessibility, Trustworthiness, and Information Quality as essential priorities. Differences arose in aspects like Novelty, Engagement, and Automation & AI, which were emphasized more by Evaluator-1.

TABLE II. INITIAL LISTS OF UX ASPECTS IDENTIFIED BY EVALUATOR-1 AND 2 USING THE MOSCOW METHOD FOR BANKING APPS.

Evaluator-1 & Evaluator-2	Evaluator-1
Usability, External Subjective Factors, Effort, Safety, Desirability, Valuable, Functionality, Aesthetic Design, External Application Assistants, Dependability, External Constraints, Independability, Updateness, Accessibility, Trustworthiness, Information Quality, Affordances, User Support, User Inclusivity, Quality Of Interaction, Context Of Use	Novelty, Automation & AI, Engagement

In the context of mobile gaming apps, Evaluator-3 and Evaluator-4 created initial prioritized lists, as presented in TABLE *III*. Shared critical aspects included Usability, Emotional Engagement, Effort, Functionality, Aesthetic Design, etc. Differences became apparent as Evaluator-3 focused on External Subjective Factors and Stimulation, etc. At the same time, Evaluator-4 prioritized the Context of Use and External Application Assistants.

TABLE III. INITIAL LISTS OF UX ASPECTS IDENTIFIED BY EVALUATOR-3 AND 4 USING THE MOSCOW METHOD FOR GAMING APPS.

Evaluator-3 & Evaluator-4	Evaluator-3	Evaluator-4
Usability, Emotional, Engagement, Effort, Functionality, Aesthetic Design, Novelty, Updateness, Accessibility, Affordances, Emerging Technologies, Automation & AI, Quality Of Interaction	External Subjective Factors, Stimulation, Desirability, Safety, User Support, User Inclusivity	External Application Assistants, Context Of Use

Similarly, TABLE *IV* presents the UX aspects identified by Evaluator-5 and Evaluator-6 for mobile health apps. Both evaluators commonly prioritized Usability, Effort, Safety, Accessibility, Trustworthiness, and Information Quality. However, they differed on aspects such as Emotional, Engagement, External Subjective Factors, Dependability, and Automation & AI. These differences highlight varied professional perspectives and considerations.

TABLE IV: INITIAL LISTS OF UX ASPECTS IDENTIFIED BY EVALUATOR-5 AND 6 USING THE MOSCOW METHOD FOR HEALTH APPS.

Evaluator-5 & Evaluator-6	Evaluator-5	Evaluator-6
Usability, Effort, Desirability, Valuable, Functionality, Safety, Updateness, Accessibility, Trustworthiness, Information Quality, User Support, User Inclusivity, Quality Of Interaction, Context Of Use	Emotional, External Subjective Factors, Engagement, Dependability, External Constraints, Independability	Aesthetic Design, Automation & AI

Although evaluators ranked aspects differently across the three domains, aligning their lists reveals that they share a common understanding of key UX aspects. These differences likely stem from the evaluators' backgrounds, skills, and available resources, highlighting distinctions between UX researchers and practitioners.

2) AHP Method Results

In banking domain [2], Evaluator-1 and Evaluator-2 both considered 'User Profiles' (UP) and 'App Domain Specificity' (ADS) more important than 'Task Complexity' (TC) and 'Feasibility' (F), as shown in TABLE *V*. Evaluator-1 rated UP much higher than TC (value = 4) and F (value = 5), while Evaluator-2 rated UP only slightly higher than TC (2) and somewhat higher than F (4). The evaluators differed in their top priority: Evaluator-1 slightly preferred UP over ADS (ratio = 2), suggesting a greater focus on user needs, whereas Evaluator-2 favored ADS over UP (ratio = 0.5), indicating a preference for aligning UX with core banking functions. Despite this contrast, both evaluators consistently prioritized user-centric and domain-specific criteria. Furthermore, both rated ADS clearly above TC and F (ADS vs. TC = 3 or 5; ADS vs. F = 4).

TABLE V. EVALUATOR-1 (E-1) AND EVALUATOR-2 (E-2) PAIRWISE COMPARISON MATRIX FOR THE CRITERIA (MOBILE BANKING).

	UP		ADC		TC		F	
	E-1	*E-2*	*E-1*	*E-2*	*E-1*	*E-2*	*E-1*	*E-2*
UP	1		2	0.50	4	2	5	4
ADC	0.5	2	1		5	3	4	4
TC	0.25	0.5	0.2	0.33	1		3	4
F	0.2	0.25	0.25	0.25	0.33	0.25	1	

In the gaming domain, Evaluators 3 and 4 agreed that UP and ADS were the top priorities, ranking them above TC and F (see TABLE *VI*). Both rated UP slightly higher than ADS (UP vs. ADS = 2), emphasizing the role of player characteristics. However, they diverged on the lower criteria. Evaluator-3 used the hierarchy: UP > ADS > TC > F, while Evaluator-4 used UP > ADS > F > TC. Both considered ADS moderately more important than TC and F (values of 2–3), but Evaluator-4 notably placed Feasibility ahead of Task Complexity (F vs. TC = 3), suggesting a focus on implementability. Overall, both valued user and domain alignment but differed on task vs. effort tradeoffs within game development.

TABLE VI. EVALUATOR-3 (E-3) AND EVALUATOR-4 (E-4) PAIRWISE COMPARISON MATRIX FOR THE CRITERIA (MOBILE GAMING).

	UP		ADC		TC		F	
	E-3	*E-4*	*E-3*	*E-4*	*E-3*	*E-4*	*E-3*	*E-4*
UP	1		2	2	2	3	3	3
ADC	0.5	0.5	1		2	3	3	3
TC	0.5	0.33	0.5	0.33	1		3	0.33
F	0.33	0.33	0.33	0.33	0.33	3	1	

In the health domain, as shown in *TABLE VII* Evaluators 5 and 6 also prioritized UP and ADS but with contrasting emphasis (see *TABLE VII*). Evaluator-5 ranked them highest (UP > ADS > TC > F), placing strong weight on user needs (UP > F = 6). In contrast, Evaluator-6 prioritized Feasibility even above User Profiles (F > UP = 3), resulting in the hierarchy: ADS > F > UP > TC. This contrast reflects differing perspectives—Evaluator-5 emphasized user-centeredness, while Evaluator-6 focused on implementation constraints, shaped by their domain experience

[2] An example of the mobile banking application case study for Evaluator-1, including all results of the UX Aspects Ranking Approach (MOSCOW and AHP), is provided in this link (https://github.com/PapersRepository/Framework-1-2-/blob/3744881ac467b48b35198178eea5713c73f7bb92/Aspects%20selection%20form%20(Evaluator-1).xlsx)

, backgrounds, areas of expertise, and roles, which influence their judgments.

TABLE VII: EVALUATOR-5 (E-5) AND EVALUATOR-6 (E-6) PAIRWISE PAIRWISE COMPARISON MATRIX FOR THE CRITERIA (MOBILE HEALTH).

	UP		ADC		TC		F	
	E-5	***E-6***	***E-5***	***E-6***	***E-5***	***E-6***	***E-5***	***E-6***
UP	1		2	0.5	4	3	6	0.33
ADC	0.5	2	1		3	5	5	0.5
TC	0.25	0.33	0.33	0.2	1		3	0.2
F	0.17	3	0.20	2	0.33	5	1	

As shown in TABLE *VIII* , all Consistency Ratios (CRs) were below the accepted 0.1 threshold, confirming high internal consistency across evaluators. CRs ranged from 0.022 to 0.085 in banking, gaming, and health, suggesting reliable pairwise judgments. This consistency strengthens confidence in the validity of the AHP-based prioritizations and their domain-specific patterns.

TABLE VIII. CONSISTENCY RATIO (CR)[3] FOR EVALUATORS' PAIRWISE COMPARISON MATRIX (IN ALL DOMAINS) FOR THE CRITERIA IN TABLE I.

Banking		Gaming		Health	
Evaluator	***CR***	***Evaluator***	***CR***	***Evaluator***	***CR***
1	0.0848	3	0.0449	5	0.0291
2	0.0530	4	0.0797	6	0.0219

a) Final UX Aspect Rankings and Evaluators' Agreement

As shown in Fig. *3*, both banking domain evaluators agreed on the critical importance of Safety (Rank 1) and Trustworthiness (Rank 3). Differences appeared in secondary aspects: User Support was ranked 2nd by Evaluator-1 but 5th by Evaluator-2, while Information Quality was ranked 2nd by Evaluator-2 but 5th by Evaluator-1. Functionality was placed 4th vs. 6th. Larger gaps occurred in lower priorities, such as Affordances (6th vs. 20th) and Context of Use (10th vs. 24th). These discrepancies illustrate how expert perspectives diverge beyond the universally high-priority factors of Safety and Trustworthiness.

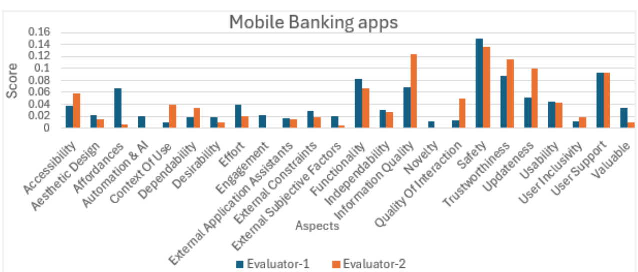

Fig. 3: The priority scores of the final UX aspects ranked list (Banking apps).

In the gaming domain (Fig. *4*), both evaluators emphasized aspects tied to entertainment and immersion. Evaluator-3 prioritized Emotional (1), Engagement (2), and Updateness (3), while Evaluator-4 ranked Engagement (1), Aesthetic Design (2), and Updateness (3). Both rated Usability highly (5th–6th) and shared focus on Emotional and Aesthetic factors, though at slightly different levels. Overall, they converged on engagement, enjoyment, and visual appeal—contrasting with the banking industry's emphasis on security and reliability.

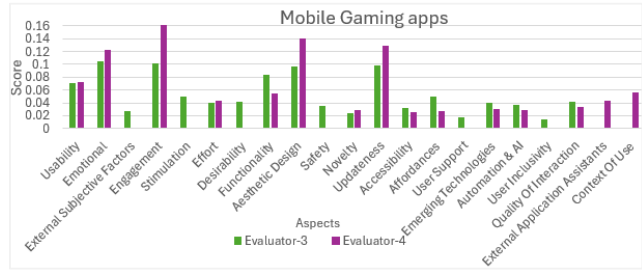

Fig. 4: The priority scores of the final UX aspects ranked list (Gaming apps).

In the health domain (Fig. *5*), Evaluator-5 prioritized Information Quality (1) and Safety (2), followed by Updateness and Accessibility. Evaluator-6 placed Functionality (1), Updateness (2), and Information Quality (3) at the top. Differences appeared in lower aspects: Evaluator-5 included Emotional (19) and Engagement (18), while Evaluator-6 emphasized Aesthetic Design (10) and Automation & AI (16). Still, both highlighted core concerns such as Accessibility, User Support, Trustworthiness, and Interaction Quality, consistent with health app priorities.

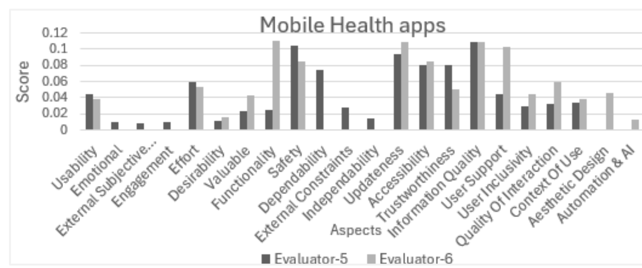

Fig. 5: The priority scores of the final UX aspects ranked list (Health apps).

b) Inter-Rater Agreement Analysis

We assessed evaluator agreement in each domain using three metrics: Kendall's Tau (rank order correlation), Top-10 Agreement (overlap of the top 10 aspects), and the Overlap Coefficient (the proportion of shared aspects relative to the list size), as shown in TABLE *IX*. For mobile banking, a Kendall's Tau of 0.12 indicates low agreement on exact rankings; however, a 50% Top-10 Agreement (5/10) and an Overlap Coefficient of 0.5 show clear consensus on core UX aspects. Despite rank differences—e.g., User Support ranked 2nd vs. 5th, and Information Quality ranked 5th vs. 2nd—both evaluators highlighted Safety and Trustworthiness as top priorities.

TABLE IX. INTER-RATER AGREEMENT RESULTS[4].

Domain	**Kendall's Tau**	**Top-10 Agreement**	**Overlap Coefficient**
Banking	0.12	50% (5/10)	0.5
Gaming	0.31	60% (6/10)	0.6
Health	0.385	70% (7/10)	0.70

[3] Note. Consistency Ratio uses Saaty's test: compute the principal eigenvalue λmax of the 4×4 criteria matrix; $CI=(\lambda max-4)/3$, $RI=0.90$ (for $n=4$), and $CR=CI/RI$. λmax per evaluator: Evaluator-1 = 4.229, Evaluator-2 = 4.143, Evaluator-3 = 4.121, Evaluator-4 = 4.215, Evaluator-5 = 4.079, Evaluator-6 = 4.059.

[4] Note. Kendall'sTau is computed on the full ranked lists. "Top-10 Agreement" is ($|Top10A \cap Top10B|/10$), where (Top10A) and (Top10B) are the two evaluators' top-10 aspect sets. The "Overlap Coefficient" is: ($|A \cap B|/\min(|A|,|B|)$), where (A) and (B) denote those same top-10 sets (both size 10 here), so it equals the Top-10 Agreement proportion.

In the gaming domain, agreement improved: Kendall's Tau was 0.31, with a 60% Top-10 Agreement (6/10) and an Overlap Coefficient of 0.6, indicating convergence on Engagement, Emotional Factors, Usability, and Aesthetics. The health domain showed the strongest consensus, with Kendall's Tau = 0.385, Top-10 Agreement = 70% (7/10), and Overlap Coefficient = 0.70, reflecting moderate rank-order consistency and clear alignment on Information Quality, Safety, Updateness, and Accessibility. Ultimately, these results indicate moderate to strong agreement in the health domain. We also computed value-based agreement using the Intraclass Correlation Coefficient (ICC) on AHP weights to assess agreement in magnitude rather than rank. TABLE *X* shows a good ICC(2,1) for Banking (0.768) and Gaming (0.805), and moderate agreement for Health (0.547). This contrast is expected: ICC captures weight magnitude, while Kendall's τ captures rank order—so high Top-10 overlap can still coincide with lower ICC values when evaluators weight the same aspects differently. When agreement is low (e.g., Kendall's $\tau < 0.20$ or Top-10 overlap < 50%), an optional Delphi-style round can be used: evaluators review anonymized rationales and aggregated rankings, then re-rate only high-disagreement items. This transparent process reconciles differences without overriding individual judgments, thereby strengthening inter-rater reliability diagnostics.

TABLE X: INTRACLASS CORRELATION (ICC) OF AHP ASPECT WEIGHTS BY DOMAIN[5].

Domain	Evaluators (pair)	Aspects (n)	ICC(2,1)
Banking	Evaluator-1 vs. Evaluator-2	21	0.768
Gaming	Evaluator-3 vs. Evaluator-4	13	0.805
Health	Evaluator-5 vs. Evaluator-6	14	0.547

In summary, although evaluators differed in their exact rankings, the metrics reveal strong consensus on key UX clusters—such as security and trust in banking, usability and engagement in gaming, and safety, reliability, and accessibility in the health domain. This highlights domain-specific priorities and shows that, despite individual differences, a shared understanding of core UX requirements exists within each application type.

B. *Results of UX Evaluation Methods Prioritization Approach*

Further, the Use Value-Effort/Complexity matrix was employed to rank UX evaluation methods adopted from [25] for the Safety aspect (highly ranked UX aspect by the evaluators) as an example, facilitating the decision-making process based on a balance between the utility of each method and its associated effort or complexity, using the criteria in Section III.B. Among the methods, as shown in Fig. *6*, "Crash reports analysis" achieved the highest priority(Upper Left Quadrant: high-value, low-effort), with a Use Value-Effort/Complexity ratio of 1.5, indicating its robust utility and relatively low implementation complexity. Similarly, "Regulatory compliance/safety standards inspections" also ranked highly (Priority 2), demonstrating its critical role in ensuring safety while requiring manageable effort (1.44). Automated evaluation, which focuses on tracking user errors, was ranked third with a ratio of 1.33, offering a reliable method that balances value and effort. Conversely, methods such as "Safety-focused heuristics" and "User testing (Error-prone task scenarios)" were assigned lower priorities (9 and 10), mainly due to their lower Use Value-Effort/Complexity ratios, which indicate a higher complexity relative to their utility. The example matrix highlights the importance of selecting methods that provide valuable insights into safety while remaining feasible within time, resource, and complexity constraints. By integrating these rankings into the development process, organizations can make informed decisions about which methods to prioritize reasonably quickly.

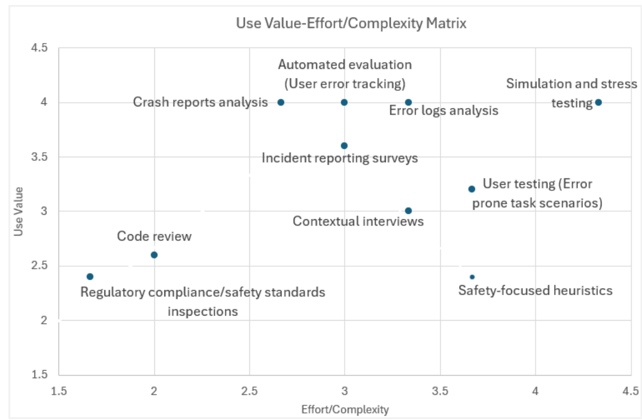

Fig. 6: The Use Value-Effort/Complexity Matrix of the Safety aspect evaluation methods[6].

V. DISCUSSION AND IMPLICATIONS

The case study results clearly demonstrate the framework's flexibility and effectiveness across various mobile app domains, including banking, gaming, and health. While UX priorities varied by domain, some key aspects remained consistent. In mobile banking, both evaluators prioritized Safety, Trustworthiness, and User Support—reflecting user expectations and financial regulations. Prior studies support these findings: Trustworthiness builds user confidence [26], Safety reduces security concerns [27] [28], and strong User Support increases satisfaction and retention [29]. Differences in lower-ranked aspects, such as Information Quality and User Inclusivity, likely stem from the evaluators' different backgrounds. As Alshammare et al. [30] noted, organizational context and experience influence UX priorities. In the gaming domain, evaluators highlighted Emotional Engagement, Aesthetic Design, Usability, and Updateness. These aspects contribute to immersive and interactive UX. Although minor differences appeared in their rankings, both evaluators agreed on the central role of engagement-related elements. The framework effectively captured these domain-specific preferences. Literature confirms that the UX in gaming is shaped by cognitive and emotional demands—including effort, frustration, and time investment—which directly impact enjoyment and satisfaction [25] [31][32][33]. In the health domain, both evaluators agreed on the importance of Information Quality, Safety, Accessibility, and Updateness -aspects critical not only

[5] Note. ICC(2,1) is a two-way random-effects, absolute-agreement, single-measure coefficient. Computed on the intersection of aspects rated by both evaluators within each domain using the AHP weights (see Fig. *3*,Fig. *4*,Fig. *5*).

[6] The computation of the Use Value–Effort/Complexity Matrix for the Safety aspect (Fig. *6*): (https://github.com/PapersRepository/Framework-1-2-/blob/e53ce98ed6d4854df75684ff129b549855ae55a7/Methods%20selection%20(Safety%20aspect%20evaluation%20methods%20ranking).xlsx).

to UX but also to patient outcomes and regulatory compliance in mHealth apps- which alignes with the Mobile App Rating Scale (MARS), standard for evaluating mHealth app quality, that emphasizes engagement, functionality, aesthetics, and information quality as core evaluation criteria [34]. The highest-ranked aspects by evaluators—Information Quality and Functionality—closely align with these MARS dimensions, further validating their prioritization. Inter-rater agreement metrics affirm the framework's reliability across domains. For example, in the health domain, Kendall's Tau reached 0.385, with a Top-10 Agreement of 70% and an Overlap Coefficient of 0.70—indicating strong evaluator consensus on core UX aspects. ICC further supported this: Banking and Gaming showed high agreement (0.768 and 0.805, respectively), while Health reflected moderate consistency (0.547). These differences are expected in expert-led evaluations but did not prevent convergence on high-priority UX elements. Across all domains, evaluators consistently identified top-ranked aspects, resulting in reliable shortlists tailored to domain-specific needs and resource constraints. Future versions of the framework could incorporate consensus-building strategies, such as Delphi rounds or group decision-making processes, to further minimize evaluator bias and resolve discrepancies in a transparent manner. The two-stage pipeline increases efficiency. The MOSCOW method quickly eliminates lower-priority aspects, then AHP compares the remaining items, lowering cognitive load and maintaining a thorough process. Unlike using either method alone, this combined approach avoids ad hoc choices and produces a clear, context-appropriate order. As shown in (Fig. *3*, Fig. *4*, Fig. *5*), the framework enables teams to select a robust top-K list with confidence. Integrating the Use Value–Effort/Complexity matrix elevates the framework from planning to decisive action. It empowers practitioners to select evaluation methods that deliver maximum results with minimal effort, positioning the framework as an indispensable decision-support tool in both academic and industrial contexts. The framework domain-independent structure, grounded in four key criteria (User Profiles, App Domain Specificity, Task Complexity, and Feasibility), promotes adaptability and generalizability. By reducing ad hoc decision-making [11], the framework enhances transparency. As technologies and user needs evolve, such a structured approach becomes increasingly valuable. To apply prioritized UX aspects effectively, organizations should align them with business objectives. They should also integrate systematic user feedback to support continuous improvement. This user-centered approach promotes relevance, inclusivity, and adaptability. Embedding the framework into agile UX cycles or ongoing design validation processes can further enhance its impact in real-world development. In conclusion, the proposed framework offers a structured, adaptable, and efficient approach for prioritizing UX aspects and selecting evaluation methods. It delivers practical benefits across mobile app development and broader digital design contexts.

The framework presents several key implications for both practice and research: **(1) Practical Efficiency:** The framework streamlines UX prioritization, helping organizations allocate resources effectively and reduce evaluation time. Its structured approach improves result reliability and supports cost-efficient, evidence-driven decisions in mobile app development. **(2) Value to Academia and Research:** This framework represents a notable advancement in UX research by offering a systematic, data-driven approach to prioritizing UX aspects. It provides a strong foundation for future studies, encouraging the exploration of new methodologies and tools for UX evaluation across a wide range of domains. Its domain-independent structure allows researchers to adapt it to diverse fields, opening opportunities to examine how UX priorities shift across different sectors, user profiles, and technological ecosystems. Additionally, the framework's integration of advanced techniques, such as AHP with real-time feedback, may offer a novel way to dynamically assess UX aspects in evolving environments, leading to more adaptive UX solutions and providing researchers with opportunities to investigate the relationship between real-time user behavior, feedback loops, and UX outcomes. Furthermore, it creates new pathways for interdisciplinary research, enabling UX to be studied alongside other fields, such as data science and artificial intelligence. By refining and extending this framework, researchers can explore how large-scale datasets or AI models, such as predictive analytics, could be leveraged to anticipate UX needs, enhance personalization, and improve user engagement. **(3) Broad Applicability:** By promoting a user-centered and flexible design process, the framework adapts to various domains and project contexts. Its scalability and structure help raise industry standards while supporting innovation and responsiveness to changing user needs.

VI. Threats to Validity

Internal validity was addressed by minimizing methodological errors and ensuring the accuracy of the results. To reduce potential measurement issues in the MOSCOW method, AHP, and the Use Value–Effort/Complexity matrix, we developed an automated Excel tool with clear instructions and examples to guide evaluators and ensure consistent calculations. Nonetheless, expert judgment remains susceptible to rater effects, such as anchoring and scale use, which we mitigated by monitoring agreement using Kendall's τ and top-10 overlap. **External validity** concerns the generalizability of the findings. Although the user experience (UX) aspects and evaluation methods were selected through a systematic mapping study [25], differences across application domains, user types, and contexts may influence broader applicability. To address this, the inclusion of three domains—banking, gaming, and health—and six evaluators provides some diversity, yet may still constrain generalizability and the assertion of domain independence. Furthermore, the use of experienced domain experts as evaluators partially addresses this limitation. Looking ahead, future studies will expand domain coverage and include user-level validation in addition to expert assessment. **Construct validity** refers to whether key factors were accurately measured. The criteria we used—outlined in Table *I* and in the Use Value–Effort/Complexity matrix—were drawn from the literature and refined with expert input to ensure meaningful weighting. To reduce misinterpretation between closely related criteria (e.g., Task Complexity vs. Feasibility), the tool includes clarifying examples. Finally, the **validity of the conclusion** was strengthened by automating all data collection and analysis steps using the tools introduced in Section III, thereby minimizing bias and increasing reliability. However, we acknowledge the absence of a comparative baseline. Future work will benchmark our hybrid pipeline against pure MOSCOW, pure AHP without filtering, and a simple weighted-sum baseline to assess the relative value and efficiency of each approach.

VII. Conclusion and Future Work

This paper presents a domain-independent framework that combines the MOSCOW method and the AHP to prioritize UX aspects in mobile apps systematically. The framework evaluates these aspects based on User Profiles, App Domain Specificity, Task Complexity, and Feasibility. It utilizes a Use Value-Effort/Complexity matrix to guide the selection of evaluation methods to enhance app adoption. The framework was validated through a case study involving three domains: mobile banking, mobile gaming, and mobile health.. The results demonstrate its effectiveness and adaptability. The framework is a valuable tool for UX professionals who must prioritize UX aspects and choose appropriate evaluation methods across various mobile app domains. Future research will focus on refining protocols for consistency, streamlining the initial list of aspects, and integrating broader expert and real-time user feedback to enhance the overall effectiveness of the system. There are also plans to expand validation efforts across additional domains. A comparative evaluation with other prioritization methods (e.g., Kano, Fuzzy, and CV) will also be conducted to contextualize the relative effectiveness. Enhancements may include leveraging historical data or large language models (LLMs) and improving the associated automated tool through collaboration with industry partners, further increasing the framework's practical utility for UX practitioners.

References

[1] A. Alslaity, B. Suruliraj, O. Oyebode, J. Fowles, D. Steeves, and R. Orji, “Mobile Applications for Health and Wellness: A Systematic Review,” Jun. 17, 2022, Association for Computing Machinery. doi: 10.1145/3534525.

[2] P. Palanque, “Ten Objectives and Ten Rules for Designing Automations in Interaction Techniques, User Interfaces and Interactive Systems,” in ACM International Conference Proceeding Series, Association for Computing Machinery, Sep. 2020. doi: 10.1145/3399715.3400872.

[3] M. Takafoli, S. Li, and V. Mäkelä, “Generative AI in User Experience Design and Research: How Do UX Practitioners, Teams, and Companies Use GenAI in Industry?,” in Proceedings of the 2024 ACM Designing Interactive Systems Conference, DIS 2024, Association for Computing Machinery, Inc, Jul. 2024, pp. 1579–1593. doi: 10.1145/3643834.3660720.

[4] M. Lee, J. Jun, S. Lee, and S. Lee, “Understanding the Initial Journey of UX Designers Toward Sustainable Interaction Design A Focus on Digital Infrastructure Energy Reduction,” in Proceedings of the 2024 ACM Designing Interactive Systems Conference, DIS 2024, Association for Computing Machinery, Inc, Jul. 2024, pp. 3079–3096. doi: 10.1145/3643834.3661598.

[5] Walter T. Nakamura, Edson Cesar de Oliveira, Elaine H.T. de Oliveira, David Redmiles, and Tayana Conte, “What factors affect the UX in mobile apps? A systematic mapping study on the analysis of app store reviews,” Journal of Systems and Software, vol. 193, 2022.

[6] E. L. C. Law, V. Roto, M. Hassenzahl, A. P. O. S. Vermeeren, and J. Kort, “Understanding, scoping and defining user experience: A survey approach,” Conference on Human Factors in Computing Systems - Proceedings, pp. 719–728, 2009, doi: 10.1145/1518701.1518813.

[7] Andy Cockburn, Philip Quinn, and Carl Gutwin, “The effects of interaction sequencing on user experience and preference,” Int J Hum Comput Stud, vol. 108, 2017, doi: 10.1016/j.ijhcs.2017.07.005.

[8] M. Hassenzahl and N. Tractinsky, “User experience - A research agenda,” Behaviour and Information Technology, vol. 25, no. 2, pp. 91–97, Mar. 2006, doi: 10.1080/01449290500330331.

[9] “Categorizing UX aspects for voice user interfaces using the kano model”.

[10] H. B. Santoso and M. Schrepp, “Importance of user experience aspects for different software product categories,” in Communications in Computer and Information Science, Springer Verlag, 2018, doi: 10.1007/978-981-13-1628-9_21.

[11] M. S. Jahan, F. Azam, M. W. Anwar, A. Amjad, and K. Ayub, “A Novel Approach for Software Requirement Prioritization,” in Proceedings - 2019 7th International Conference in Software Engineering Research and Innovation, CONISOFT 2019, Institute of Electrical and Electronics Engineers Inc., Oct. 2019, pp. 1–7. doi: 10.1109/CONISOFT.2019.00012.

[12] A. Dhouib, A. Trabelsi, C. Kolski, and M. Neji, “Prioritizing the Usability Criteria of Adaptive User Interfaces of Information Systems based on ISO/IEC 25040 Standard,” vol. 22, no. 4, pp. 107–128, 2017, doi: 10.3166/ISI.22.4.107-128ï.

[13] J. Tang, “Prioritizing User Requirements based on Analytic Hierarchy Process,” in Journal of Physics: Conference Series, IOP Publishing Ltd, Apr. 2021. doi: 10.1088/1742-6596/1861/1/012021.

[14] M. Pallot et al., “Evaluating user experience as a means to reveal the potential adoption of innovative ideas,” IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC), pp. 1–11, 2020.

[15] E. Castillejo, A. Almeida, and D. López-de-Ipiña, “Modelling users, context and devices for adaptive user interface systems,” 2014, Emerald Group Publishing Ltd. doi: 10.1108/IJPCC-09-2013-0028.

[16] L. E. Gutiérrez et al., “Combined Framework of Multicriteria Methods to Identify Quality Attributes in Augmented Reality Applications,” Mathematics, vol. 11, no. 13, Jul. 2023, doi: 10.3390/math11132834.

[17] J. Deutschländer, A. C. Weigand, A. M. Klein, D. Winter, and M. Rauschenberger, “There Are no Major Age Effects for UX Aspects of Voice User Interfaces Using the Kano Categorization,” in International Conference on Web Information Systems and Technologies, WEBIST - Proceedings, Science and Technology Publications, Lda, 2023, pp. 330–339. doi: 10.5220/0012187600003584.

[18] C. Malinka, U. Von Jan, and U. V. Albrecht, “Prioritization of Quality Principles for Health Apps Using the Kano Model: Survey Study,” JMIR Mhealth Uhealth, vol. 10, no. 1, Jan. 2022, doi: 10.2196/26563.

[19] Klingbeil, M and Renz, J and Bruechner, and D, Prioritization of user experience aspects for learning environments. IATED, 2020.

[20] Rajeshkumar, S and Omar, Ridha and Mahmud, and Murni, “Taxonomies of user experience (UX) evaluation methods,” international conference on research and innovation in information systems (icriis), pp. 533–538, 2013.

[21] M. Schrepp et al., “On the Importance of UX Quality Aspects for Different Product Categories,” International Journal of Interactive Multimedia and Artificial Intelligence, vol. 8, no. 2, pp. 232–246, 2023, doi: 10.9781/ijimai.2023.03.001.

[22] F. Sitorus, J. J. Cilliers, and P. R. Brito-Parada, “Multi-criteria decision making for the choice problem in mining and mineral processing: Applications and trends,” May 01, 2019, Elsevier Ltd. doi: 10.1016/j.eswa.2018.12.001.

[23] A. M. Klein, J. Deutschländer, K. Kölln, M. Rauschenberger, and M. J. Escalona, “Exploring the context of use for voice user interfaces: Toward context-dependent user experience quality testing,” Journal of Software: Evolution and Process, vol. 36, no. 7, Jul. 2024, doi: 10.1002/smr.2618.

[24] E. Eshet and H. Bouwman, “Addressing the Context of Use in Mobile Computing: A Survey on the State of the Practice,” Nov. 23, 2012, Oxford University Press. doi: 10.1093/iwc/iwu002.

[25] Alshammare H, Alshayeb M, and Baslyman M, “Revealing the Mobile UX Horizon: Exploring User Experience Aspects, Attributes, and Measurement Methods-A Systematic Mapping Study,” Comput Stand Interfaces, 2025.

[26] K. Gokmenoglu, MKorhan K and Amir, and Aysel, “The impact of perceived fairness and trustworthiness on customer trust within the banking sector,” Journal of Relationship Marketing, vol. 20, pp. 241–260, 2021, Accessed: Jan. 03, 2025.

[27] O. E. Adediran, “An Evaluation of Usability and Security in Mobile Banking,” 2024, doi: 10.21203/rs.3.rs-4881972/v1.

[28] Mathur, Tanuj and VV, and Ravi Kumar, “Human-Computer Interaction (HCI) in Banking: Users' Website Usability Experience and Self-Confidence,” International Seminar on Application for Technology of Information and Communication (iSemantic), pp. 174–177, 2024, Accessed: Apr. 26, 2025.

[29] M. Muzaffarov, “Miasto Przyszłości Kielce 2023 DIGITALIZATION OF THE BANKING SECTOR, ITS TRANSFORMATION AND EFFICIENCY,” 2023.

[30] H. Al-Shammare, M. Alshayeb, and M. Baslyman, “Towards a Readiness Model for Usable-Software Development in Organizations,” Softw Pract Exp, 2024, doi: 10.1002/spe.3388.

[31] M. Csikszentmihalyi and M. Csikzentmihaly, Flow: The psychology of optimal experience. 1990.

[32] K. Karpouzis and G. Yannakakis, Emotion in Games. 2016.

[33] I. Granic, A. Lobel, R. E.-A. psychologist, and undefined 2014, “The benefits of playing video games.,” psycnet.apa.orgPaperpileI Granic, A Lobel, RCME EngelsAmerican psychologist, 2014•psycnet.apa.orgPaperpile, 2013.

[34] Y. Terhorst et al., “Validation of the Mobile Application Rating Scale (MARS),” PLoS One, vol. 15, no. 11 November, Nov. 2020, doi: 10.1371/JOURNAL.PONE.0241480