A Novel Framework for Usability Evaluation of Mass-based Software

Harshita Sharma
Department of Computer Science and Engineering
Netaji Subhas University of Technology
New Delhi, India
harshita.sharma.phd23@nsut.ac.in
ORCID: 0009-0009-5496-9363

Ritu Sibal
Department of Computer Science and Engineering
Netaji Subhas University of Technology
New Delhi, India
rsb@nsut.ac.in
ORCID: 0009-0008-5982-8925

Pinaki Chakraborty
Department of Computer Science and Engineering
Netaji Subhas University of Technology
New Delhi, India
pinaki.chakraborty@nsut.ac.in
ORCID: 0000-0002-2010-8022

***Abstract*—Mass-based software platforms such as instant messaging apps, video conferencing platforms, and generative AI interfaces play a pivotal role in modern e-businesses. Many businesses use these software as an outreach for new customers. Their usability, hence significantly affects digital engagement, customer satisfaction, and customer retention. In view of the limitations of the existing literature, tools and techniques for mass-based software, this research work proposes a four phase methodology with the intent to develop and design tools, techniques and a framework for the empirical evaluation of various performance and human-factor related parameters like effectiveness, efficiency, satisfaction, cognitive load, etc. widely used in the domain of usability evlauation. In addition, this research work also aims to propose and automate usability heuristics adapted especially to the use of mass-based software for the context of e-businesses. For e-business platforms, the framework provides actionable insights into UX design, supports detection of usability issues, and facilitates informed design decisions for improved customer retention in today's customer-centric digital transformation.**

Keywords—usability testing, mass-user systems, measured data, self-reported data, human factors, e-business, user experience.

I. INTRODUCTION

A. *Usability Evaluation*

Usability evaluation is the process of systematically evaluating the usability aspects of a product. Users take good usability of a product for granted and do not explicitly appreciate usability when it is proper. Users are concerned about usability only when they have trouble in using a product and, at times, switch over to a competing product with better usability [1]. One of the primary aspects of usability evaluation is usability testing which commonly involves some members of the target audience to use the product and assess their interaction with the product. Usability testing may be performed for debugging, i.e. before launching the product, or quality assessment, i.e. after the product has been deployed. The importance of usability testing lies in the fact that it simulates how users interact with the product in the real world and provides a realistic assessment of the user interaction. In the context of e-business, usability directly influences user engagement, which in turn can provide insights regarding customer satisfaction, and ultimately, business outcomes such as conversion rates and revenue. Digital platforms rely heavily on pleasant user experiences to retain customers and maintain their competitive edge in today's close cut competition with similar products. Even minor usability issues can lead to abandoned transactions or poor customer reviews, making systematic usability testing an essential practice in the digital commerce landscape.

B. *Mass-based Software*

Mass-based software refers to widely used digital applications designed for large-scale, heterogeneous user populations, typically accessed across varied platforms, devices, and contexts. Communication-centric platforms such as WhatsApp and Zoom, social networking sites like Instagram and Facebook, UPI-based financial apps, and AI-powered tools such as ChatGPT and Copilot are some examples of such software. A key distinguishing factor between mass-based software and other types of software like context specific and professional software is that the other types of software is usually used only by a particular group of people working in the relevant domain, whereas mass-based software is used by casual users with generic capabilities spanning millions of people around the world, and also across varying types of devices and network conditions. These are popular among a vast and diverse range of population across the globe. Their functionalities can range from providing day-to-day communication, video conferencing, information retrieval to entertainment. Given the wide range of people with varying capabilities, challenges, expectations and preferences, designing, developing, evolving and testing this software pose unique challenges. Usability testing for such mass-based software becomes even more crucial as the user experience impacts not only a small group of users, but millions of people.

Although the exact term "mass-based software" has not been used by researchers, usability of generic software has been explored by some researchers [2]. This research introduces the term "mass-based software" to formalize this category and address the need for dedicated usability frameworks tailored to the challenges posed by high-scale, general-purpose applications in modern digital ecosystems.

C. *Usability Testing Parameters*

Various parameters can be used to test the usability a product. They can be categorized as performance related parameters and human-factor related parameters. Performance related parameters provide ways to objectively assess the various usability aspects of a product and provide a strong empirical foundation of a usability study. These include parameters like efficiency and effectiveness measured using metrics such as task completion time, number of errors and task success rate. In e-business platforms, these performance metrics hold a significant amount of importance. For instance, prolonged task completion time during a checkout process or frequent user errors in navigating a payment gateway can significantly impact the customer's decision to complete a transaction. Such issues not only affect user satisfaction but also lead to direct business losses, highlighting the economic relevance of usability testing in commercial digital environments. In

the context of mass-based software, performance-related parameters become especially important due to the scale and diversity of users involved. These are usually evaluated on the basis of metrics derived from measured data which refers to quantifiable data that can be captured/recorded in numeric values based on the user's interaction with a product.

In contrast to empirically observed performance related parameters, human-factor related parameters assess the physical and mental state of the users while using a software product. For example, cognitive load, satisfaction and affinity are subjective parameters based inherently on the individual user's capabilities, mental and emotional states before, after and during the task(s). They are typically assessed using psychological scales [3-4]. Unlike performance-related metrics, which focus on what users do, human-factor parameters emphasize how users feel and think during and after the interaction. Evaluating human-factor parameters is especially crucial for mass-based software, where user retention and satisfaction can significantly impact success and can help to uncover subtle issues with accessibility and engagement which impacts user preferences. These aspects are especially important for e-business services, where user trust, comfort, and emotional response can determine whether a customer returns to the platform again, abandons it or recommends the platform to a friend. Human-factor related parameters usually rely on self-reported data which is data that the users themselves provide based on their experience, emotions, perceptions and thoughts during and after using a product, either through structured or unstructured mechanisms. The most widely used tool for collecting self-reported data is by the use of standard questionnaires and surveys including the use of both closed and open-ended questions. Self-reported data can also be conducted by various types of interviews. In contrast to measured data, self-reported data is subjective in nature. However, this type of data can easily be biased and not-so reliable being influenced by the mood, expectations, etc., of the diverse users. For this reason, self-reported data is most valuable when interpreted in conjunction with objectively measured data.

D. *Heuristic Evaluation*

In the context of usability testing, heuristic evaluation are practical and general guiding principles based on patterns of known human/user behavior that can guide developers, designers and evaluators to identify, address and work on various usability issues [5]. Heuristic evaluation is a usability inspection method where a number of experts in the given domain come together to assess a user interface against a set of established usability principles, called heuristics, even before the product is tested with a set of representative users to unveil usability issues. Each expert gives his individual evaluation against a given set of heuristics, which the designers combine to get a comprehensive understanding of the various usability issues.

E. *Significance for e-businesses*

Mass-based software, widely used in the e-business domain, includes instant messaging apps, video conferencing platforms, social networking sites, UPI payment apps, Generative AI tools and the like. These platforms are designed to serve large and diverse user bases, often spanning different regions, languages, and levels of digital literacy. The usability of such systems has a direct impact on user engagement and business performance, as delays, confusion, or friction in interactions will lead to customer drop-off. As such, usability testing plays a crucial role in identifying barriers to user success and ensuring that digital services meet the needs of all users, not just a narrow segment.

In the context of e-business and digital services, mass-based software forms the backbone of customer interaction, support, and engagement. Minor usability issues can cascade into widespread dissatisfaction, reduced customer retention, and ultimately economic loss. Hence, the usability of mass-based platforms needs to be evaluated using both performance metrics and human-factor indicators that are scalable, reliable, and adaptable to heterogeneous user profiles. In e-business it is essential to set high standards for the customer perceived usability, reliability and performance of your solution [6].

This paper, based on a currently ongoing doctoral thesis work, aims to present a comprehensive overview of the significance of usability testing of mass-based software for e-business platforms as well as the frameworks to be developed and the phase-wise methodology used. This research aims to develop a systematic and empirically grounded framework for the usability evaluation of mass-based software, particularly in the context of e-business platforms. The work intends to investigate both performance-related and human-factor related usability parameters through measured and self-reported data, propose novel automated heuristics adapted for mass-user systems, and design tools and techniques to support scalable, reliable usability assessments. The work aims to study four categories of mass-based software namely communication-centric platforms like instant messaging apps (WhatsApp, Telegram, etc.) and video conferencing platforms (Zoom, Google Meet, etc.) , social networking sites (Instagram, etc.), Unified Payment Interface (UPI) apps (Gpay, Paytm, etc.), and generative AI tools (ChatGPT, Copilot, etc.).

II. Research Problem

Mass-based software, used by millions globally like instant messaging apps, video conferencing platforms, social networking sites, generative AI tools, and the like play a key role in today's accelerating digital transformation. With the capability to support large-scale user interactions, virtual collaboration, AI-assisted services, and customer engagement across geographies and devices, various mass-based software is becoming central to the modern e-business infrastructure. The usability testing of such mass-based software, hence is of prime importance. In addition to facilitating diverse user groups around the world for their day-to-day tasks, mass-based software also provide leverage to e-businesses by enabling their reach to a wide variety of customers and acting as an operational backbone for user interaction. In e-business environments, usability directly influences key business metrics such as conversion rates, customer satisfaction, digital service adoption, and platform retention. Poor usability can increase abandonment rates, raise support costs, and reduce the effectiveness of digital engagement strategies. Thus, robust usability testing methods not only improve interface design but also enhance overall digital business performance. They also serve as a repository of generating user-experience related data like engagement time, abandonment rates, etc. This data when mined and

analyzed, can bring out useful insights to business owners regarding their shortcomings and consequent improvements required.

However, there is a lack of a comprehensive framework and research thereof for usability evaluation of mass-based software. Usability testing is based on the evaluation of various performance based parameters, human-factor related parameters and usability heuristics. Most current usability evaluation studies rely only on subjective input for human-factor related parameters. Techniques for empirical observance and objective assessment of these parameters would generate more reliable results and insights of these factors in usability evaluation studies. Existing usability heuristics are limited in their applicability to mass-based software, especially for use in the context of e-businesses. Existing heuristics were developed with general-purpose software in mind and often assume a relatively uniform user group, a controlled environment, and a limited scale of usage. Also they do not address e-business related factors like trust, privacy and the like. Automation of heuristic evaluation would also be a welcome development, as it will reduce the dependency on experts.

This research aims to develop a formal model, software tools and heuristics to objectively assess performance and human-factor related parameters of usability and use them to perform usability testing of mass-based software.

III. Outline of Objectives

The primary objective of this research work is to enable e-businesses to assess and enhance the usability of mass-based software platforms more objectively—to improve digital customer journeys, operational scalability, and user-centric product strategies. This will provide e-businesses with quantifiable data rather than simply subjective data. This can yield more concrete insights regarding how e-businesses can use mass-based software to increase the reach of their products. This comprehensive goal is intended to be achieved via fulfilling four specific objectives as follows:

RO1: To develop a comprehensive usability model for mass-based software critical to digital businesses, focusing on performance and human-factor parameters.

RO2: To design techniques and software tools for the objective measurement of human-factor usability aspects in applications widely used in e-business contexts, supporting data-driven user experience improvements.

RO3: To formulate and automate heuristic evaluation methods specifically adapted for mass-based software that are integral to digital business ecosystems.

RO4: To empirically test the developed usability model and tools across different categories of mass-based software, generating actionable insights for enhancing the user-centric design of digital services in e-business environments.

IV. State of the Art

The idea of usability stems from the concept of being user-friendly, and is one of the most important quality attributes of a software and can be stated as simply being the "ease of use of the software". [7] The IEEE 610.12-1990 standard defines usability as "the ease with which a user can learn to operate, prepare inputs for, and interpret outputs of a system or component" [8]. Despite the varying features proposed and used by researchers over the years to study usability, three attributes are the most widely used namely effectiveness, efficiency and satisfaction. [9]. The ISO 9241-11 standard also defines these three as the key attributes of usability [10]. It defines effectiveness in terms of "the accuracy and completeness with which users achieve their specified goals", efficiency as "the resources expended in relation to the accuracy and completeness with which users achieve goals" and satisfaction as "the comfort and acceptability of use". Learnability is another important usability aspect, especially crucial for mass-based software. Given that most users of these software are casual users; the learnability curve should not be steep. Neilson, a pioneer in the field of usability relates learnability to how quickly a user is able to start getting work done from the system [11].

Instant messaging apps used multiple times on a day-to-day basis by millions of people globally are one of the top candidates for mass-based software. Studies have been carried out on various instant messaging apps like WhatsApp, Telegram, WeChat, Facebook Messenger and the like to assess their usability quotient [12-15]. The usability parameters of effectiveness, efficiency, satisfaction, and learnability have been the focus of the majority of the research work done on these apps. Video conferencing platforms like Zoom, Google Meet, Microsoft Teams, that facilitate synchronous communication by transmitting audio, video, and other forms of data between users over networked environments are also another category of mass-based software that have seen rapid growth in view of the increased reliance on remote and hybrid modes of interaction. Given their central role in digital communication, several researchers have ventured into this domain with accessibility, performance under bandwidth constraints, privacy and security being the focus of usability studies.[16-19]. Social networking sites like Instagram, Facebook etc. also constitute an important category of mass-based software that allows people to connect and interact with thousands through a single post. These have also caught researchers' attention working in the usability domain who have evaluated these sites for user engagement, interface design, and requirement elicitation from user-generated content [20-25]. Information retrieval and generative AI tools like ChatGPT, Microsoft's Copilot and similar applications also present a new domain of study for usability researchers with the most widely studied aspects being interaction quality, and task effectiveness [26-28].

In regards to performance related parameters, most research studies have kept their focus on three parameters namely effectiveness, efficiency and learnability [29-32]. Effectiveness is usually measured as a binary success metric that ignores task complexity and partial success. Introduction of task weighting can help present a more accurate view of this parameter. Efficiency assumes equal time cost across tasks, ignoring the fact that some tasks are inherently simpler as compared to other more complex tasks. This shortcoming can be overcome by normalization. Learnability is measured over multiple controlled sessions. For human-factors related parameters, satisfaction is the most widely evaluated parameter measured via popular standard questionnaires like the System Usability Scale (SUS) [33] and the Post-Study System Usability Questionnaire (PSSUQ) [34]. Cognitive load is another important human-factors related parameter with great relevance to the usability of mass-based software.

Popular standard tools like the NASA Task Load Index (TLX) [35] questionnaire and the Subjective Mental Effort Questionnaire (SMEQ) [36] are commonly used. However, a major shortcoming of these tools is their subjective nature and dependence on user recall and perception which can be highly biased. Research has also been done on mass-based software using usability heuristics. These works have focused on usability issues affecting user performance, engagement, and accessibility [37-38]. Studies on generative AI tools however reveal the need for clearer feedback, error recovery, more customization options, improved transparency and context awareness as prime usability issues [39]. However, a major drawback of existing studies is that they are largely expert-based, there is no automation or log-based evaluation and there is a heavy reliance on visual UI reviews. Also no existing heuristics have been adapted or automated for usability testing of mass-based software.

Based on the literature review, four prominent research gaps have emerged. Firstly, there is no comprehensive usability model for mass-based software integrating both performance and human-factor related parameters. Second, there are no standard equations for measuring performance related parameters. Thirdly, in a majority of the literature, human-factor related parameters are assessed using self-reported data and the techniques are inaccurate. Lastly, there are no usability heuristics defined for mass-based software.

V. Methodology

This doctoral research work will be carried out in a multi-phase strategy using a mixed-methods approach focused around the four research objectives. The first phase focuses on usability studies on communication-centric mass-based software like instant messaging apps and video conferencing platforms. The second phase will be centered around the development of a prototype tool that logs behavioral data and correlates these with the results of human-factor related parameters obtained via subjective assessment. It is intended to develop a tool that can track user parameters like eye gaze, emotions, click patterns, no-action time during tasks and co-relate and compare the results obtained using these metrics with those obtained via subjective assessment of human-factor parameters. The third phase will be focused on the development and automation of heuristics for mass-based software. Currently, software is evaluated by experts against generalized heuristics. It is intended to adapt and develop heuristics especially catered to be used for usability evaluation of mass-based software. Automation of these heuristics is also planned to reduce the dependency on the availability of experts. The final phase will test the usability of different types of mass-based software using the model and tools developed in this study. A high-level conceptual block design diagram is presented in fig 1.

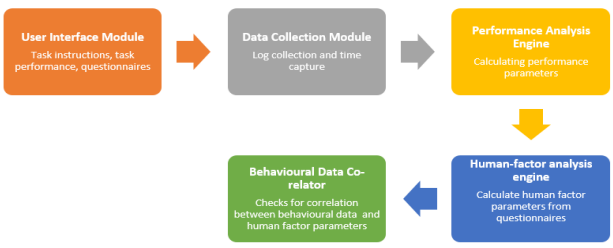

Fig. 1. High-level conceptual block design of the proposed tool

Data will be gathered via in-lab experiments conducted with participants, preferably with their own devices due to one's familiarity and comfort level with own device. A sample size of 30 participants for such experimental studies in usability evaluation is considered acceptable. It is intended to keep the sample uniform by recruiting participants from undergraduate computer science courses, which ensures comparable technical capabilities between participants. Observations and recording of data will be done either by manual observation by the authors in-person or via recording tools like UXtweak for performance related metrics. For subjective assessment, standard questionnaires disseminated via google forms to the participants will be used. Data analysis will be carried out statistically using relevant statistical software available. By combining technical objectivity with user-centric analysis, the proposed methodology supports digital businesses in identifying user experience gaps, improving service delivery, and improving customer reach across platforms.

To ensure the reliability, applicability, and suitability of the proposed model, a multi-dimensional validation approach will be adopted. Firstly, a theoretical validation will be carried out by consultation and reviews by scholars and practitioners in the field of usability. This will ensure conceptual soundness and completeness. This will be followed by an empirical validation by carrying out controlled usability studies involving end users performing real-world tasks. Both performance metrics (e.g., success rate, error rate, task time) and human-factor metrics (e.g., SUS, NASA-TLX scores) will be collected and analyzed. This two-tier validation strategy aims to ensure that the model is not only theoretically robust but also empirically grounded and generalizable across real-world contexts.

VI. Expected Outcome and Contributions

The proposed work can address a pressing need of the hour for both academics and industry. Mass-based software being used by millions globally, can potentially impact e-businesses worldwide. From an academic point of view, this research can contribute not only to the field of user experience and usability engineering but also provide insights for the discipline of cognitive ergonomics. This model is expected to provide a basis for future empirical studies and comparative evaluations. This research work can have a significant impact on the field of usability studies by devising novel techniques for objective evaluation of human-factor related parameters. Two major contributions of this work that is found lacking in most current usability research studies is the development of tools and techniques that can capture and correlate the human-factor parameters to observable and measurable behavioural data. This can help in producing more reliable results as compared to the current practice of using only subjective assessment techniques for collecting human-factors related data. This in turn can give focused directions and areas of improvement to work upon for various e-businesses. The vision of proposing and automating heuristics for mass-based software will be another novel contribution of this work. It can offer a new, cost-effective and time-saving direction in usability testing both for academia and industry. This would not only reduce the dependence on human experts but also provide a more reliable, more efficient and repeatable methodology of heuristic evaluation.

The demonstration of usage of the developed ideas and solutions in this study for mass-based software can provide ideas to future researchers to tailor and customize these for other types of software as well. For industry, the proposed tools and models developed in this study can help reduce dependency on lab-based testing and enable usability assessment in remote, real-world settings. By offering objective insights based on user behavior these tools can help product teams detect usability bottlenecks earlier in the development cycle, saving time and reducing the cost of redesign based on user testing usually done much later in the product life cycle. The research contributes to the strategic goals of digital businesses by enabling better product-market fit, increased user retention, and more efficient digital experience design—critical drivers in today's digital economy.

VII. Stage of the Research

So far, phase one centered on usability studies on communication-centric mass-based software like instant messaging apps and video conferencing platforms has been completed. From the experiment on instant messaging apps where we compared WhatsApp, Facebook Messenger, Snapchat and Telegram on three performance parameters namely effectiveness, efficiency and learnability and two human-factor parameters namely satisfaction and a novel parameter affinity. We found that WhatsApp has the best effectiveness. WhatsApp, Telegram and Facebook Messenger have comparable efficiency. WhatsApp is also better than the other apps in terms of satisfaction and affinity. To evaluate learnability, the experiment was split over two sessions. We found that there was no major improvement in effectiveness and efficiency during the intervention period. This low learnability may be because the interface of instant messaging apps is designed to facilitate exploration by users and not to optimize performance. From the experiment on video conferencing platforms, where we compared Zoom, Google Meet, Microsoft Teams, Cisco Webex and Skype, it was observed that Skype had the best performance (effectiveness and efficiency) amongst all the five video conferencing apps, while Cisco Webex had the poorest performance. Microsoft Teams, although lagging in performance, outperformed others in terms of effects on users, with respect to satisfaction and cognitive workload. The focus will now be on developing a tool that logs behavioral data and correlate these with the results of human-factor related parameters obtained via subjective assessment. It is intended to complete development of this tool till December 2025. Fig. 2 depicts the current stage of this research work. Adapting and developing usability heuristics for mass-based software in the context of e-businesses as well testing of these techniques on mass-based software will be the agenda of the first half of the upcoming year.

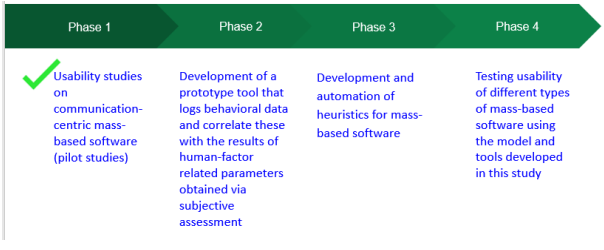

Fig. 2. Proposed phase-wise strategy

VIII. Conclusion

This doctoral work proposes a structured approach for usability evaluation of mass-based software by combining performance-related metrics with self-reported human-factor data. The work proposes to fill the existing gap of measurement of human-factor data using observable behavioral data as compared to self-reported subjective data prevalent across current research. It also proposes to fill another major gap of customized usability heuristics adapted for mass-based software and their automation. Given the increasing reliance on large-scale digital platforms and the dependence of e-businesses on these platforms, the proposed framework addresses a critical need for empirical usability testing of mass-based software.

References

[1] M. L. Tan, R. Prasanna, K. Stock, E. E. Doyle, G. Leonard, and D. Johnston, "Usability factors influencing the continuance intention of disaster apps: A mixed-methods study," *Int. J. Disaster Risk Reduct.*, vol. 50, Art. no. 101874, 2020.

[2] M. Li and P. Nielsen, "Making usable generic software: A matter of global or local design?," in *Proc. 10th Scandinavian Conf. Inf. Syst. (SCIS)*, 2019.

[3] T. Kosch, J. Karolus, J. Zagermann, H. Reiterer, A. Schmidt, and P. W. Woźniak, "A survey on measuring cognitive workload in human-computer interaction," *ACM Comput. Surv.*, vol. 55, no. 13s, Art. no. 244, pp. 1–38, Dec. 2023.

[4] M. A. Faudzi, Z. C. Cob, M. Ghazali, R. Omar, and S. A. Sharudin, "User interface design in mobile learning applications: Developing and evaluating a questionnaire for measuring learners' extraneous cognitive load," *Heliyon*, vol. 10, no. 18, Art. no. e37494, 2024.

[5] A. Nistico, D. Markudova, M. Trevisan, M. Meo, and G. Carofiglio, "A comparative study of RTC applications," in *2020 IEEE International Symposium on Multimedia (ISM)*, Dec. 2020, pp. 1–8.

[6] A. Rudolf and R. Pirker, "E-Business testing: user perceptions and performance issues," in *Proc. First Asia-Pacific Conf. Quality Software*, Hong Kong, China, 2000, pp. 315–323.

[7] K. Sagar and A. Saha, "A systematic review of software usability studies," *Int. J. Inf. Technol.*, vol. 9, no. 4, pp. 1–24, Dec. 2017.

[8] IEEE, *IEEE Standard Glossary of Software Engineering Terminology*, IEEE Std 610.12-1990, Dec. 1990.

[9] P. Weichbroth, "Usability of mobile applications: A systematic literature study," *IEEE Access*, vol. 8, pp. 55563–55577, 2020.

[10] International Organization for Standardization, *Ergonomics of human-system interaction – Part 11: Usability: Definitions and concepts*, Standard 9241-11, 2018.

[11] J. Nielsen, *Usability Engineering*. San Francisco, CA, USA: Morgan Kaufmann, 1994.

[12] S. Caro-Álvaro, E. García-López, A. García-Cabot, and A. Mavri, "Measuring the effects of usability recommendations for mobile instant messaging apps on user performance and user behaviour," *Int. J. Hum.-Comput. Interact.*, vol. 41, no. 1, pp. 608–626, 2025.

[13] S. Caro-Álvaro, E. García-López, A. García-Cabot, L. de-Marcos, and A. Domínguez-Díaz, "Applying usability recommendations when developing mobile instant messaging applications," *IET Softw.*, vol. 16, no. 1, pp. 73–93, 2022.

[14] T. N. Wen, P. L. Cheng, and P. L. Chang, "Evaluate the usability of the mobile instant messaging software in the elderly," in *MEDINFO 2017: Precision Healthcare Through Informatics*, Gothenburg, Sweden, 2017, pp. 818–822.

[15] S. Caro-Álvaro, E. García-López, A. García-Cabot, L. de-Marcos, and J. J. Martínez-Herraiz, "Identifying usability issues in instant messaging apps on iOS and Android platforms," *Mob. Inf. Syst.*, vol. 2018, Art. no. 2056290, 2018.

[16] J. T. Simon-Liedtke, W. K. Bong, T. Schulz, and K. S. Fuglerud, "Remote evaluation in universal design using video conferencing systems during the COVID-19 pandemic," in *Proc. Int. Conf. Hum.-Comput. Interact.*, Cham, Switzerland: Springer, 2021, pp. 116–135.

[17] A. Grigoriu, “Evaluating remote communication applications using student usability reviews,” in *Proc. Int. Conf. Comput.-Hum. Interact. Res. Appl.*, Cham, Switzerland: Springer Nature, 2024, pp. 213–229.

[18] D. Oosthuizen, T. Sharma, J. Chavula, and M. Densmore, “Investigating the usability and quality of experience of mobile video-conferencing apps among bandwidth-constrained users in South Africa,” in *Proc. 43rd Conf. South Afr. Inst. Comput. Sci. Inf. Technol. (SAICSIT)*, Cape Town, South Africa, Jul. 18–19, 2022, EPiC Series in Computing, vol. 85, EasyChair, pp. 243–256

[19] N. Samrgandi, “Privacy, security, and usability of mobile video conferencing applications: A comparative study,” *Int. J. Adv. Appl. Sci.*, vol. 10, no. 11, pp. 109–116, Nov. 2023.

[20] A. Baj-Rogowska and M. Sikorski, “Exploring the usability and user experience of social media apps through a text mining approach,” *Eng. Manag. Prod. Serv.*, vol. 15, pp. 86–105, 2023.

[21] A. E. Agazzi, “Study of the usability of LinkedIn: A social media platform meant to connect employers and employees,” *arXiv* preprint arXiv:2006.03931, 2020.

[22] M. Sihag, *From videos to requirement: A data-driven approach for finding requirements relevant feedback from TikTok and YouTube*, Ph.D. dissertation, Univ. of Victoria, Victoria, BC, Canada, 2023.

[23] L. Bouillon, D. Schouten, and F. Campos, “Designing the user experience to achieve better engagement and feedback on TikTok accounts,” in *Proc. Int. Conf. Human-Comput. Interact.*, Cham, Switzerland: Springer Nature, Jul. 2023, pp. 503–515.

[24] F. Abdillah, “Evaluation analysis of the use of the TikTok application using the System Usability Scale (SUS) method,” *J. Vocat. Educ. Inf. Technol.*, vol. 4, no. 2, pp. 57–61, Dec. 2023.

[25] M. Sihag *et al.*, “A data-driven approach for finding requirements relevant feedback from TikTok and YouTube,” in *Proc. 31st IEEE Int. Requirements Eng. Conf. (RE)*, Hannover, Germany, Sep. 2023, pp. 111–122.

[26] G. Bubaš, A. Čižmešija, and A. Kovačić, “Development of an assessment scale for measurement of usability and user experience characteristics of Bing chat conversational AI,” *Future Internet*, vol. 16, no. 1, p. 4, 2023.

[27] N. Nikolaidis *et al.*, “A comparison of the effectiveness of ChatGPT and Co-Pilot for generating quality Python code solutions,” in *Proc. 2024 IEEE Int. Conf. Software Anal., Evol. Reeng. – Companion (SANER-C)*, Mar. 2024, pp. 93–101.

[28] A. Darejeh, N. Marcus, G. Mohammadi, and J. Sweller, “A critical analysis of cognitive load measurement methods for evaluating the usability of different types of interfaces: Guidelines and framework for human-computer interaction,” *arXiv* preprint arXiv:2402.11820, 2024.

[29] R. Harrison, D. Flood, and D. Duce, “Usability of mobile applications: Literature review and rationale for a new usability model,” *J. Interact. Sci.*, vol. 1, art. no. 1, 2013.

[30] O. Korhan and M. Ersoy, “Usability and functionality factors of the social network site application users from the perspective of uses and gratification theory,” *Qual. Quant.*, vol. 50, no. 4, pp. 1799–1816, 2016.

[31] N. Bevan, J. Carter, J. Earthy, T. Geis, and S. Harker, “New ISO standards for usability, usability reports and usability measures,” in *Lect. Notes Comput. Sci.*, vol. 9731, pp. 268–278, 2016.

[32] B. A. Kumar and P. Mohite, “Usability of mobile learning applications: A systematic literature review,” *J. Comput. Educ.*, vol. 5, no. 1, pp. 1–17, 2018.

[33] J. Brooke, “SUS—a quick and dirty usability scale,” in *Usability Evaluation in Industry*, P. W. Jordan, B. Thomas, B. A. Weerdmeester, and I. L. McClelland, Eds., London, UK: Taylor & Francis, 1996, pp. 189–194.

[34] J. R. Lewis, “Psychometric evaluation of the post-study system usability questionnaire: The PSSUQ,” in *Proc. Hum. Factors Soc. Annu. Meeting*, vol. 36, no. 16, pp. 1259–1260, Oct. 1992, Los Angeles, CA, USA: Sage Publications.

[35] S. G. Hart and L. E. Staveland, “Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research,” in P. A. Hancock and N. Meshkati, Eds., *Human Mental Workload*, Amsterdam, The Netherlands: North-Holland, 1988, pp. 139–183.

[36] F. R. H. Zijlstra and L. van Doorn, “The construction of a scale to measure subjective effort,” Technical Report, Dept. of Philosophy and Social Sciences, Delft University of Technology, Delft, The Netherlands, 1985.

[37] S. Caro-Alvaro, E. García-López, A. García-Cabot, and A. Mavri, “Measuring the effects of usability recommendations for mobile instant messaging apps on user performance and user behaviour,” *Int. J. Hum.–Comput. Interact.*, vol. 41, no. 1, pp. 608–626, 2025.

[38] W. Maulana, S. R. Natasia, and M. G. L. Putra, “UI/UX evaluation of Android-based social media Sestyc application with think aloud and heuristic method,” in *AIP Conf. Proc.*, vol. 2693, no. 1, p. 020009, Nov. 2023.

[39] A. Pascual Almenara and S. Sayago Barrantes, “Towards designing a set of usability and accessibility heuristics focused on cognitive diversity: An exploratory case study with generative artificial intelligence,” *Information*, vol. 15, no. 12, p. 769, 2024.