What is User Engagement?: A Systematic Review of 241 Research Articles in Human-Computer Interaction and Beyond

Bernard J Jansen
Qatar Computing Research Institute
Hamad Bin Khalifa University
Doha, Qatar
jjansen@acm.org

Kathleen W Guan
Division of Psychology and Language Sciences
University College London
London, United Kingdom
kathleen.guan.20@ucl.ac.uk

Joni Salminen
University of Vaasa
Vaasa, Finland
jonisalm@uwasa.fi

Kholoud Khalil Aldous
Qatar Computing Research Institute
Hamad Bin Khalifa University
Doha, Qatar
kkaldous@hbku.edu.qa

Soon-Gyo Jung
Qatar Computing Research Institute
Hamad Bin Khalifa University
Doha, Qatar
sjung@hbku.edu.qa

Abstract

User engagement (UE) is widely discussed in HCI articles, but its definition, reliability, and application remain elusive. This research conducts a systematic literature review of 241 articles from 1993 to 2023 to analyze how UE is defined and measured within the domain of HCI. Our findings reveal significant definitional inconsistencies that hinder UE's practical application in HCI research and system design. Based on our findings, we recommend using UE as a categorical label rather than a unified construct until more systematic frameworks are established. We also highlight the need for divergent views of UE across HCI research communities as a valuable avenue to pursue. This divergent view approach can help HCI researchers focus on specific, measurable aspects of UE that align with specific community practices and norms. Our findings also suggest that until such a framework emerges, researchers should be aware of its limitations when using UE as a research construct.

CCS Concepts

• **Human-centered computing**; • **Human computer interaction (HCI)**;

Keywords

user experience, user interaction, user needs, user metrics

ACM Reference Format:
Bernard J Jansen, Kathleen W Guan, Joni Salminen, Kholoud Khalil Aldous, and Soon-Gyo Jung. 2025. What is User Engagement?: A Systematic Review of 241 Research Articles in Human-Computer Interaction and Beyond. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/3706598.3713505

ACM ISBN 979-8-4007-1394-1/2025/04
https://doi.org/10.1145/3706598.3713505

1 Introduction

User engagement (UE) is frequently mentioned as a construct in human-computer interaction (HCI) research, yet its definition and practical application remain ambiguous and inconsistent. In scientific research, a *construct* refers to an abstract concept developed to represent phenomena that cannot be directly observed but can be inferred from measurable indicators [78]. Constructs serve as theoretical tools that help researchers describe and study complex phenomena, such as human behaviors [15]. In much HCI research, UE is perceived as a construct encompassing various cognitive, affective, and behavioral dimensions of user interaction with technology. While UE cannot be directly observed, it is inferred through different metrics such as time spent, user actions, and self-reported experiences [54]. This research reviews the fragmented definitions of UE, evaluates UE's use in HCI research, and proposes a change in how UE could be defined and used within HCI.

Let us first briefly consider two attempts to rigorously define a construct, one from the physical and the other from the non-physical sciences: *force* and *attention*. These are both theoretical concepts: not directly observable but reliably inferred from other measured data [80]. Many sciences, including, among others, cognitive sciences, information systems, and HCI, leverage such indirectly observable constructs [28, 56].

In Newtonian physics, *force* represents the push or pull between objects [61]. Force cannot be directly observed but is a theoretical construct developed to help make sense of object interactions, inferred from measurable elements such as mass, acceleration, velocity, and time. Although its definition is an abstraction, conceptually, force has remained unchanged since Newton's proposal in 1687. There is broad agreement and empirical support concerning its theoretical foundations, and the factors defining force are limited and straightforward to measure; there is even a specific measurement for force, a *Newton.* This clarity and consensus make force an exemplar construct.

On the other hand, *attention* is a construct that has seen extensive refinement in cognitive sciences. Attention refers to the selective salience of specific information in a given context. Despite cognitive sciences being a relatively new field, the concept of attention has undergone extensive empirical investigation, leading to increasingly refined and precise measurement methods, such as reaction-time tasks and neuroimaging [32]. Yet, the applicability of attention as a construct continues to be debated [83], such as concerning the role of selective salience in the study of attention [21]. In this article, we seek to facilitate this form of constructive, epistemic dialogue focusing on UE.

Our analysis emphasizes the HCI perspective, though the findings may also interest other disciplines. As highlighted in studies on umbrella constructs [19, 45], these constructs that attempt to encapsulate broad phenomena often suffer from definitional fragmentation, which, as we will show, is also the case with UE. This fragmentation can lead to issues in operationalization and validity, as seen with the usability construct in HCI [75, 76]. Other HCI constructs, such as interaction [22]and personas, have come under similar critical scrutiny [10, 11]. Like with these other examples, the current application of UE presents challenges in achieving consistency across HCI studies, further complicating UE's practical applicability in system design, user behavior (UB), and other HCI research. Umbrella constructs like UE require a clear, agreed-upon conceptualization to ensure their relevance and utility for HCI research and applications. *Relevance* is the degree to which UE findings align with real-world UBs and system interactions, making these findings applicable and meaningful to HCI challenges [13, 72]. *Applicability is* the practical application of these findings [35, 77], such as guiding the design of user interfaces, enhancing user experience (UX), and informing the development of interaction strategies. A well-defined UE construct would allow researchers to develop consistent methodologies and metrics, facilitating accurate comparisons across studies and improving replicability. This consistency could enhance the theoretical and practical applicability of the UE construct in HCI, supporting the creation of systems that better meet user needs and behaviors. While one may expect more fuzziness with abstract concepts relative to physical ones, if such a construct is frequently applied in scientific research and practice, one would still expect acknowledgment of the abstraction, reasonably clear and agreed-upon definition, theoretical grounding, and practical benefit. To this end, a critical examination of UE can aid the quality and progress of HCI research and practice.

The current research addresses several knowledge gaps by investigating the following research questions (RQs):

RQ1: *How is user engagement defined?*

RQ2: *How is user engagement measured?*

RQ3: *Are conceptualization and assessment of user engagement aligned across studies?*

For RQ1, we analyze and report on the definitions of UE presented in the literature, identifying the specificity and categories of UE definitions across various thematic areas of HCI research. While there is an overarching field of study in the HCI domain, within this domain, several research communities concentrate on specific topics such as accessibility, health, the web, mobile devices, and privacy. These research communities are groups of researchers who share common interests and goals within the broader HCI domain. Each community may approach concepts and methodologies differently, but they all contribute to the collective understanding and advancement of HCI research. This analysis is important because if UE is to have value as a research construct, its definitions should generally be precise and consistent across or perhaps within HCI research communities and studies. For RQ2, we analyze and report the metrics applied to measure or evaluate UE. Ideally, there ought to be some agreement on what is indirectly measured to quantify UE and whether these measures should be consistent across HCI research communities within HCI. For RQ3, we analyze and report on the alignment between UE as defined within HCI studies and how UE is measured. If definitions and measures are aligned, this might indicate UE is conceptually focused and consistently applied within a given study. If definitions and metrics are misaligned, this might suggest that the UE may be used as theoretical or model 'window dressing' for empirical research.

This article presents a literature review of prior UE reviews, followed by our research premises. We then present our methodology for our systematic literature review on UE. Subsequently, we present our literature review results, returning to our RQs. We follow this with a discussion of the implications in the HCI field. Finally, we propose a way forward for employing UE that can help understand people and their interactions with technology.

2 Foundation for Investigation and Review of Literature

2.1 Challenges of Defining UE

Various attempts have been made to define UE definitively across the HCI field, but as we will show, further progress is needed. An example of a previously proposed definition includes "a quality of user experience characterized by attributes of challenge, positive affect, endurability, aesthetic and sensory appeal, attention, feedback, variety/novelty, interactivity, and perceived user control" [54, p. 1]. There are other similar examples in the HCI UE literature, which we examine in detail below. Most of these definitional attempts are extremely broad. For example, UE has been denoted as a concept that refers to the quality of the user experience (UX) that emphasizes the positive aspects of the interaction with technology [3, 66, 71]. Lalmas et al. [37, p, 1] attempt a definition as UE "refers to the quality of the UX that emphasizes the positive aspects of interacting with an online application and, in particular, the desire to use that application longer and repeatedly."

From this lineage of UE definitions, some organizations and practitioners (including designers, product managers, and marketing managers) have adopted the UE concept as an objective, target, or goal to optimize, whether qualitatively or quantitatively (see Table 1). This maxim sees high UE as an antecedent to continued system use, often mediated by improvements in a system's UX or usability [55, 84], whereas low UE indicates an intent of discontinuation [67, 70]. However, a challenge of this view of UE is that it might not necessarily reflect genuine value for system end users [12, 53]. Others have framed UE as having a potentially negative effect or at least been associated with secondary negative effects, as identified in the HCI research topics dealing with (social media) addiction, impulsive behaviors, nudging, manipulation, and dark design patterns [5]. Often, these are inadvertent consequences of "optimizing" UE: designers are not deliberately seeking to harm the users, but harmful and adverse outcomes like addiction emerge as side effects of maximizing UE.

Table 1: Positive and negative connotations of UE based on different ways of defining it.

HCI research tries to. . .	Positive connotation	Negative connotation
maximize or optimize	•UE as a target key performance indicator (KPI) or metric •UE as a UX goal •UE as a usability goal •UE as not to be a distractor from other target KPIs or metrics •UE to fit with the organization's system design culture •E as one of a range of options, including usability and effectiveness	N/A
minimize or mitigate	N/A	•UE as addictive use of a system (e.g., social media feeds) •UE as a form of impulsive behavior, such as online purchasing or social media commenting •UE to avoid the illusion of data validity (preoccupation with metrics) •UE to avoid the dark patterns of design (i.e., a feature that subtly encourages users to perform a specific action)

Lamas and colleagues [37] note that a challenge in studying UE is the lack of an agreed-upon definition and understanding of UE and, therefore, how UE can be measured, whether it should be viewed as a desired [85] or undesired phenomenon, and should researchers aim at maximizing or minimizing UE (and under which boundary conditions)? So, the employment of UE has yet to clarify system use, and its varying definitions confuse what kind of systems we should design. Building upon the exploration of prior research and the challenges surrounding UE definitions, this study addresses these issues through a systematic review. By synthesizing existing literature and identifying patterns in definitions and measurements, we aim to provide a clearer perspective on UE's role within HCI.

2.2 Prior User Engagement Reviews

There have been prior attempts at literature reviews of UE within the HCI literature. Hochheiser and Lazar [20] propose a model of social engagement based on the influences that motivate HCI responses to societal issues and the types of interventions that professionals in these fields can contribute to. Although relevant to UE, this research focuses on broader social and related implications rather than the psychological aspects of HCI. Relatedly, Doherty and Doherty [14] systematically review how HCI research defines, understands, and measures engagement, including UE. The authors analyze 351 articles and 102 definitions of engagement from different domains and perspectives, proposing an ecological framework that distinguishes engagement as a state, an agent, or an interaction. The review examined engagement in diverse ways across the spectrum of the cross-disciplinary research community.

In contrast, other key efforts have defined UE directly in using a system, such as a study by O'Brien and Toms [54] which aims to synthesize related theories of UB and psychology like flow, play, aesthetics, and information interaction. O'Brien and Toms define UE as a quality of UX characterized by attributes of challenge, positive affect, endurability, aesthetic and sensory appeal, attention, feedback, variety/novelty, interactivity, and perceived user control, encompassing a broad spectrum of attributes. They also suggest that UE is a process that consists of four stages: (1) point of UE, (2) period of sustained UE, (3) disengagement, and (4) re-engagement; however to our knowledge these explicit stages are not empirically tested, In later work, O'Brien [50] expounded on these reportedly theoretical foundations of UE, namely the *Flow Theory* and *Philosophy of Experience*, attempting to deconstruct UE using principles for evaluating concepts of clarity, scope, and meaning. Rather than a literature review, as is our research, this more recent work is more of a commentary highlighting possibly epistemological influences on UE. Finally, O'Brien et al. [52] have also investigated measuring UE via existing scales. This previous work focused on efforts to define UE. We update and enhance these important efforts in our current work by critically comparing and contrasting definitions (including new definitions that have emerged since) to continue discussing the concept of UE as a construct.

Peters et al. [57] explore the concept of engagement in HCI, including UE, and how it relates to concepts such as attention, emotion, interest, and immersion. The authors examine different definitions and perspectives on engagement from various domains and categorize them according to the perception-cognition-action loop. They also discuss the factors and phases of UE and the challenges of measuring and modeling it. The article is an exploratory attempt to clarify the meanings and implications of UE for designing and implementing intelligent interfaces, and it can be considered more of a general scoping overview of engagement rather than a focused and systematic examination of UE, as presented in this work. Similarly, Garett et al. [16] review UE in website design, identifying elements that influence UE, like navigation, graphical representation, organization, content utility, purpose, simplicity, and readability. They note how website design is critical for UE, but there is no consensus on defining and measuring the specific elements of effective website design. This review focuses on the narrow aspect of website design and measures. In contrast, our systematic review goes beyond visual artifacts to address the range of HCI areas and dimensions of UE as presented in the HCI literature.

Sigerson and Cheng [65] focused specifically on the psychometric properties of twelve scales for measuring UE in social network sites. The researchers report mixed evidence for the validity and reliability of the scales, with some scales being more rigorously validated than others. The article also identifies limitations and biases of the scales, such as sampling and acquiescence biases. Likewise, Ng et al. [48] examined specifically on UE indicators of mental health apps, while our systematic review covers more than indicators and addresses multiple HCI areas. Ng et al. systematically review how UE indicators are measured and reported for apps, specifically mental health. They find a high heterogeneity and inconsistency in the definitions, criteria, methods, and thresholds used to assess UE indicators across studies. Finally, Suh and Cheung [69] carried out a literature review of 59 empirical studies on UE. The researchers concluded that UE can be conceptualized as a psychological state, a behavioral experience, or a combination of both. They found that UE can have different dimensions, such as cognitive, affective, behavioral, and meaningful, and UE can be affected by technology, social, content, and individual characteristics [79]. Also, UE can have various consequences, such as extended technology use, decreased social interaction, task inefficiency, and poor information technology appraisal. The article proposes a comprehensive (or perhaps all-encompassing) framework for UE that accounts for the interplay among the key elements associated with UE.

Our systematic literature review significantly builds on prior research, further scrutinizing the UE concept. Some prior reviews have proposed different models based on the literature reviews. For example, O'Brien and Toms [54] UE is a multi-stage process involving sustained interaction, disengagement, and re-engagement. Meanwhile, Doherty and Doherty's [14] ecological framework distinguishes engagement as a state, agent, or interaction; and Peters et al.'s [58] present a motivation-engagement-thriving model of UE. While there are also few efforts to evaluate UE's benefits for user understanding or system design, the limited evaluations show no to minimal positive impact [48]. Rather, past work has largely focused on proposing new definitions of UE that encompass a conglomeration of the cognitive, affective, and behavioral spectrums. Subsequently, these diverse definitions lead to calls highlighting the difficulty of implementing and measuring UE without using exceedingly complex frameworks or approaches. As such, despite extensive efforts by researchers, the concept of UE remains elusive and multidimensional. This suggests a crucial gap in identifying and clarifying the core psychological and behavioral dimensions of UE and establishing a more coherent and consistent understanding of UE within HCI. The current state of UE research highlights the need for a more concerted effort to delineate the key dimensions of UE and refine its theoretical underpinnings: *while UE is widely cited, there is no consistent agreement on its definition, metrics, or theoretical foundations.* The current work highlights the fragmented state of UE research and proposes avenues for refining UE with more empirically grounded categorical labels or narrower definitions within HCI communities. This involves synthesizing existing literature and conducting rigorous empirical research to validate and expand upon these dimensions. The following section outlines our methodology for collecting and analyzing relevant literature.

3 Methodology

3.1 Overview

A systematic search using specific search phrases related to UE (detailed further below) to address our RQs was conducted on the major article databases of ACM Digital Library (ACMDL), IEEE Xplore, and Web of Science (WoS). These databases were selected because of HCI research coverage (especially ACMDL) and extensive coverage of journal articles (WoS). Other databases like Scopus were discarded due to likely high overlap with the chosen three databases. We searched the three databases using nearly identical approaches, varying the search only for the peculiarities of the archive platforms. For an additional coverage check, we also searched the Google Scholar archives to identify articles explicitly defining UE. We limited the searches to peer-reviewed journal articles and conference proceedings published from 1993 through 2023, including all full-text articles from journals or conferences written in English. We explain the process in the following subsections.

3.2 Article Collection and Screening

Figure 1 illustrates the literature collection process using the PRISMA flow diagram [64]. PRISMA (**P**referred **R**eporting **I**tems for **S**ystematic Reviews and **M**eta-**A**nalyses) is a set of guidelines and checklists, a standard approach for reporting systematic reviews [68]. We present the PRISMA flow diagram and process [2] to report our systematic data collection process.

ACM Digital Library: We searched the ACMDL with the phrase "user engagement" (quotes included) in the article title using the advanced search bar, as the ACMDL is a comprehensive archive of HCI research arguably containing the leading HCI conferences and journals in the field. Our rationale was that articles with UE explicitly in the title would predominantly focus on UE. The ACMDL also spans various HCI fields, and user-related research is conducted. We searched with a date limit through 31 December 2023 for full-text research articles published by ACM, journals or conferences, and full papers as the filters. The search yielded 158 results published through 2023.

Web of Science: To mitigate possible source bias by collecting user studies from just the ACMDL database (despite it being a diverse archive and to HCI outlets), we searched the WoS, another comprehensive archive of user research, with the phrase "user engagement" (quotes included) as the keyword in the article title using the advanced search feature. The search yielded 570 results published through 2023.

IEEE Xplore: We also search the IEEE Xplore database to mitigate possible source bias. However, manual screening of the titles and abstract did not locate relevant articles, so no articles from IEEE Xplore were included in our systematic review.

Google Scholar: Although the searches on ACMDL and WoS databases yielded many relevant articles concerning UE, many articles did not explicitly define UE. We used the opportunity to

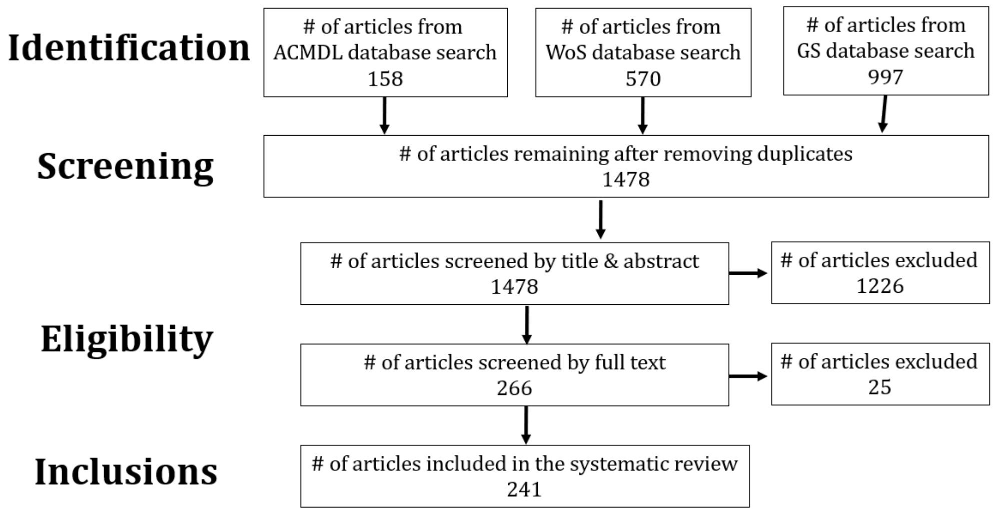

Figure 1: PRISMA flow diagram of the literature collection process, with a full explanation provided in the text.

expand our systematic review to a third archive, Google Scholar (GS), as a modified snowballing technique similar to the approach in [62]. This substantially increased the coverage of the literature (n = 997). Wohlin [81] points out that different techniques should be employed to maximize the chances of identifying relevant literature. Using the *Publish or Perish* [18] software, we searched GS for the phrase "user engagement," with a date limit through 2023, for full-text research articles.

Overall Process: The combined searches on archival databases and our snowball sampling resulted in 1,478 articles after removing duplicate articles. We first manually screened the articles by inspecting whether each retrieved article reported relevant UE research in the title and abstract, and then, for those that passed this screening, we screened the full text for the remaining articles. We followed a defined protocol to determine inclusion and exclusion criteria for our systematic review. Articles were included if they (1) explicitly referred to UE within the context of HCI, (2) provided empirical data or theoretical analysis related to UE, and (3) were published in peer-reviewed journals or conference proceedings between 1993 and 2023. Articles were excluded if they (1) discussed UE in non-technological or purely theoretical settings without application to HCI, (2) focused on engagement without a technological aspect, or (3) were non-English publications. Examples of exclusion were UE involving customers and non-technology products and other forms of engagement not with a piece of technology. Applying the inclusion/exclusion criteria via manual inspection led to 1237 articles (86.7%) from the three datasets that did not meet our inclusion/exclusion criteria from either the title, abstracts, or full text being discarded. The pertinent information was recorded in a spreadsheet for each selected article, which we included as supplementary material.

Overall, using multiple databases, the additional coverage check, and the rigorous screening process led us to believe that our review covers a major portion of UE research in the HCI context. We chose ACMDL and WoS due to their extensive indexing of high-impact journals and conference proceedings in HCI and technology-related fields. GS supplemented the specialized databases by capturing a broad spectrum of literature, conference proceedings, and emerging works that may not be indexed in the other databases, thereby enhancing comprehensiveness without compromising the HCI-specific focus of the primary databases used. Therefore, the current papers cover major HCI outlets, including conferences and journals.

3.3 Data Collection and Methods of Analysis

We extracted a range of information from the articles, as shown in Table 2.

Manual coding was completed by reading the articles and recording the information in a spreadsheet. We defined article attributes based on our open coding, using standard accepted definitions of terms as labels (see Table 2). The article title, publication type, and year of publication were straightforward. The definition of UE, location of definition, data collection, analysis type, and how engagement was measured were also relatively straightforward, as these were expressed directly in the article. Major takeaways and additional comments were qualitative expressions by coders. A team of three researchers participated in the review process for the analysis. Each researcher played a specific role: one primary coder conducted the initial open coding of all 241 articles, a secondary coder independently verified twenty percent of the data for interrater reliability, and a third researcher oversaw the thematic categorization and resolved any discrepancies. This collaborative approach ensured consistency and reliability in analyzing UE definitions and metrics. Cohen's Kappa statistic was used to quantify the agreement, accounting for the possibility of chance agreement. The obtained Kappa value of 0.8 indicated substantial agreement between the coders.

Concerning the definition analysis, one of the authors independently coded all the articles. The process of coding definitions involved a word-by-word analysis of each definition, with explicit characteristics coded (i.e., tallied) based on the presence of relevant keywords per definition. As definitions of UE in our sample were

Table 2: The attributes extracted from the articles. [A] denotes automatic retrieval using ACMDL, WoS, and GS export functions, while [M] denotes manual extraction by the researchers.

Attribute	Automation [A]/ Manual [M]	Definition
Article title	A	The name of the article. This is directly the article.
Publication type	A / M	This is from the article meta-data. Most are journals or conferences. This is directly from the article reference.
Year of publication	A	This is from an article reference. This is directly from the article.
Thematic area	M	The research community of the article. Examples include e-commerce, health, education, social media, music, etc.
Definition of engagement	M	This is how the article defines UE (if defined). This should be a quote.
Location of definition	M	Page number where UE defined
Data collection	M	If an empirical study, how was the data collected? Examples would be user studies, surveys, or user logs.
Analysis type	M	The methods employed in the article. Examples would be quantitative (i.e., ANOVA, t-test, etc.), qualitative (i.e., grounded theory, open coding, etc., mixed (i.e., both quantitative and qualitative), or algorithmic (i.e., machine learning).
How engagement measured	M	What metrics are used? Examples would be page views, likes, shares, comments, etc.
Major takeaways	M	The major findings or implications.
Additional comments	M	Interesting aspects of the article or highlights that make the research interesting.

diverse, many definitions had several characteristics (and their categories) coded. The extraction of UE definitions was limited to seventy percent (n=181) of the 241 articles explicitly defining UE. Articles that did not include definitions (n=60) were excluded from this analysis portion. The process involved manually reviewing the full text of each article to identify and record definitions. This selective approach allowed us to concentrate on articles that significantly contributed to understanding the variability in UE definitions.

We used multiple methods to examine the UE definitions. First, we used similarity analysis [82] to determine how closely the definitions aligned with one another, relying on Google's Universal Sentence Encoder [9] to calculate semantic similarity scores between each pair of definitions. This helped quantify the alignment or divergence in how UE is defined across the literature. The similarity score ranged from 0 (not similar) to 1 (identical). The similarity scores allowed us to identify patterns and degrees of definitional overlap. Linguistic term analysis [30] was conducted to identify the most frequently occurring terms across definitions, offering insights into common conceptual elements. These analyses are grounded in natural language processing (NLP) techniques, which are widely used to understand and compare textual content across various domains [25, 33].

Finally, we applied thematic analysis [17] to categorize definitional trends into broader themes. This multi-method approach allowed us to systematically capture the diversity and patterns in the UE literature. Thematic areas were developed using an open-coding technique [17], which is appropriate when there is a need for emergent taxonomies [23]. A thematic area refers to a defined sphere of knowledge. This taxonomy was developed inductively by analyzing the types of studies in the sample. The major thematic areas identified were (1) social media, (2) health, (3) education, (4) information technology, (5) e-commerce, and (6) software, though other thematic areas also emerged (see Table 3).

4 Results

4.1 Exploratory Findings

We first present exploratory findings, beginning with an overview of the UE articles as presented in Table 3.

Outlets and Year of Publication: Table 3 shows that 132 articles (57.6%) were published in journals, while 97 (42.4%) were conference articles. Figure 2 shows the publication of UE articles by year. Data shows that researchers were increasingly using the UE, at least beginning in 2013 through 2020. This increased usage does not alleviate the lack of mutually agreed upon definitions. Instead, the situation worsens as new operational definitions and understandings of UE keep emerging, thus resulting in the conceptual fragmentation of the concept.

Thematic areas: Regarding thematic areas, the most common area was generic (i.e., no specific research community), with 58 articles (23.8%), followed by social media with 33 articles (13.5%), health with 22 articles (9.0%), and education with 21 articles (8.6%). There were also articles related to IT, e-commerce, software technology, gaming, crowdsourcing, news, entertainment, sports, media, tourism, product design, politics, art, humanities, non-profit, manufacturing, energy industry, accessibility, apparel, consumer goods, enterprise resource planning, finance, hospitality, mobile phone retail, security, and sustainability. This variation in thematic areas illustrates the pervasive emergence of UE across research fields.

Data Collection: The most used method for data collection was surveying in 105 studies (34.9%), followed by user logs and user studies in 66 studies (21.9%) each. Interviews were used in 31 studies (10.3%), experiments in 23 studies (7.6%), reviews in 6 studies

Table 3: Overview of User Engagement Articles, including publication types, thematic areas, data collection methods, and analysis methods used in 241 research articles on UE.

Attribute	N	%
Publication Type	**N**	**% of Articles**
Journal Articles	139	57.6
Conference Articles	102	42.4
Total	241	100.0
Thematic Areas (articles may have more than one)	**N**	**% of N**
Generic	58	23.8%
Social Media	33	13.5%
Health	22	9.0%
Education	21	8.6%
Information Technology	15	6.1%
E-commerce	10	4.1%
Software Technology	14	5.7%
Gaming	7	2.9%
Crowdsourcing	6	2.5%
News	6	2.5%
Entertainment	5	2.0%
Sports	5	2.0%
Media	7	2.9%
Tourism	4	1.6%
Product Design	3	1.2%
Politics	3	1.2%
Art	2	0.8%
Humanities	2	0.8%
Non-Profit	2	0.8%
Manufacturing	2	0.8%
Energy Industry	2	0.8%
Accessibility, Apparel, Consumer Goods, Energy, Enterprise Resource Planning, Finance, Hospitality, Mobile Phone Retail, Security, Sustainability (1 each)	15	6.1%
Total	244	100.00
Data Collection (articles may have more than one)	**N**	**%**
Surveys	105	34.9%
User Log	66	21.9%
User Study	63	20.9%
Interviews	31	10.3%
Experiment	23	7.6%
Review	6	2.0%
Case Study	3	1.0%
Experience Sampling, Simulation, In-Situ (1 each)	4	1.3%
Total	301	100.0%
Analytics Approaches	**N**	**% of Articles**
Quantitative	176	73.0%
Qualitative	21	8.7%
Mixed Methods	36	14.9%
N/A (non-empirical article)	8	3.3%
Total	241	100.0%

(2.6%), and case studies in 3 studies (1.0%). Some articles used experience sampling, simulation, and *in-situ* data collection methods. This variation in data collection methods illustrates the selective employment of UE across different research designs. The relatively high representation of survey data highlights the operationalization of UE as a self-reported measure. The findings also demonstrate how reported research on UE is conducted across diverse thematic areas and employs various data collection and analysis methods. Based on these exploratory analyses, this diversity in thematic areas, data collection methods, and analysis approaches indicates

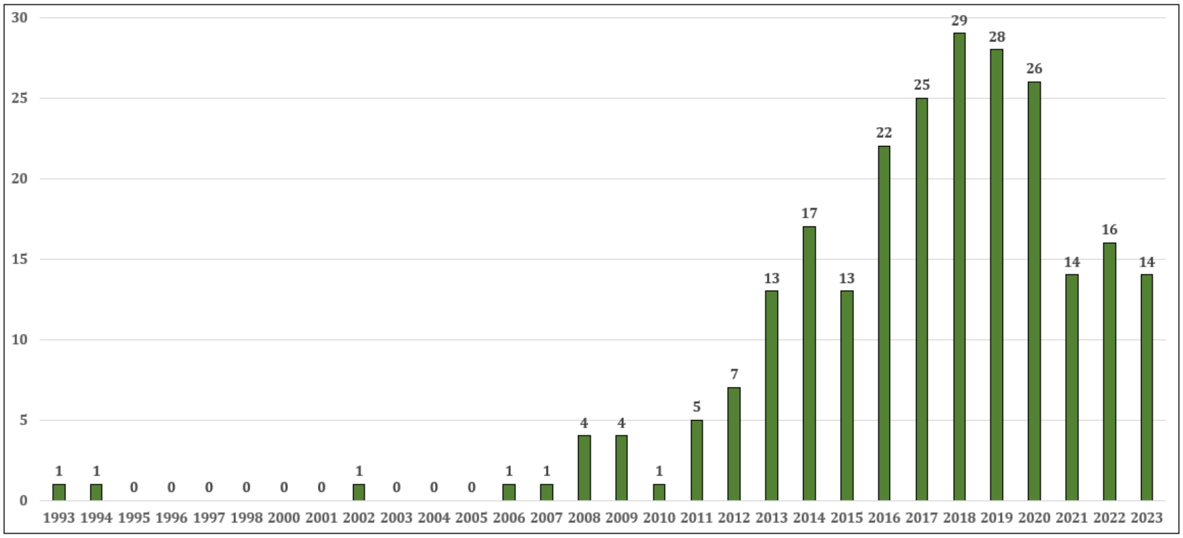

Figure 2: UE publications by year. Researchers are increasingly using the UE concept, at least through 2020, even though it needs clear conceptual underpinnings. This increased usage does not address the lack of mutually agreed-upon definitions. Note: Our search identified no UE articles from 1995 to 2001 and 2003 to 2005.

issues concerning consistent definitions and metrics of UE and standardized data collection and analysis. From these exploratory results, we now move to address our RQs.

4.2 RQ1: How Is User Engagement Defined?

We extracted the definitions from 181 (75.1%) of the 241 articles for the UE definitions. We next examine the characteristics of 181 definitions of UE, beginning with definition length, linguistics, and similarity analysis. The length of a UE definition ranges between 13 to 487 characters and 2 ("salient intermediary") to 72 words after removing stop words. The average number of characters per definition is 110 characters (SD = 72.7), and the average number of words per definition is 16 terms (SD = 10.7). We combine the 181 definitions of UE into one string variable. We also enabled the removal of the default stop word (that is, commonly used words). The total number of unique words and bigrams (two words) is 535. The total sum of the term frequency (TF) (how often a term occurs) of the 535 words is 1,580 terms.

The ten most frequently used words are user (TF = 94, 5.9% of total terms), interaction (TF = 45, 2.8%), user experience (TF = 34, 2.2%), quality (TF = 32, 2.0%), system (TF = 25, 1.6%), time (TF = 25, 1.6%), involvement (TF = 18, 1.1%), use (TF = 15, 0.9%), cognitive (TF = 14, 0.9%), and attention (TF = 14, 0.9%). Table 4 shows the top ten words and bigrams used in the 181 UE definitions. We checked the similarity between each pair of the 181 definitions. Using the Universal Sentence Encoder [9], we encoded all the definitions and calculated the semantic similarity [82] (based on the likeness of their meaning or semantic content). Table 4 shows the top 10 unigrams and top 10 bigrams used in the 181 UE definitions. Figure 3 shows the histogram of the semantic similarity scores and the count of definition pairs for each semantic similarity score.

Here, we have 181 definitions, which form 16,290 unique pairs of definitions. Thus, we have 16,290 semantic similarity scores for the definition pairs. The histogram in Figure 3 shows that 4,970 out of 16,290 (30.5%) of the definition pairs have similarity scores between 0.1 and 0.2, and 3,338 (20.5%) have less than 0.1 similarities. Most of the pairs of definitions, 15,601 (95.8%), have a similarity score of less than 0.5, indicating a significant variation in how researchers define UE. The rest of the 689 (4.2%) pairs have a similarity score greater than or equal to 0.5, suggesting some degree of agreement or similarity in these definitions.

These linguistic and semantic results (95.8% having a similarity score of less than 0.5) indicate a broad spectrum of perspectives and approaches to defining UE. The relatively low occurrence of the frequency of keywords (most frequent keyword occurring 5.9%) in the UE definitions and the single-digit occurrences of UE definitions with high similarity (4.2% having a similarity score greater than or equal to 0.5) already indicate a wide diversity in UE definitions. Most definition pairs have low similarity scores, indicating a need for more consistency in understanding and interpreting UE within HCI research.

To evaluate if there is variance within research communities, we also calculated the definition's semantic similarity by focusing on the research community having the highest number of definitions, 26, which is social media. For the social media area, there are 324 unique pairs of definitions. The histogram in Figure 4 shows that 82 of 324 (25.3%) of the definition pairs have similarity scores between 0.2 and 0.3, and 76 (23.5%) have similarities between 0.3 and 0.4. Most of the pairs of definitions, 258 (79.6%), have a similarity score of less than 0.5, indicating a significant variation in how researchers define UE even in the same community. The rest of the pairs, 66 pairs (20.4%), have a similarity score greater than or equal to 0.5, suggesting some degree of agreement or similarity in these definitions. From this analysis, the within-community definitions have a

Table 4: The ten most frequent words in the UE definitions with TF and the percentage of the total number of words (% of total words). The ten most frequent bigram words in the UE definitions with TF and the percentage of the total number of words (% of total words).

Word (unigram)	TF	% of total words	Bigram	TF	% of total words
user	94	5.9%	user experience	34	2.2%
interaction	45	2.8%	emotional cognitive	13	0.8%
quality	32	2.0%	positive aspect	11	0.7%
system	25	1.6%	novelty interactivity	8	0.5%
time	25	1.6%	feedback novelty	8	0.5%
involvement	18	1.1%	behavioral connection	8	0.5%
use	15	0.9%	sensory appeal	7	0.4%
cognitive	14	0.9%	perceived control	6	0.4%
attention	14	0.9%	awareness motivation	6	0.4%
technology	14	0.9%	motivation interest	6	0.4%
			technological resource	6	0.4%

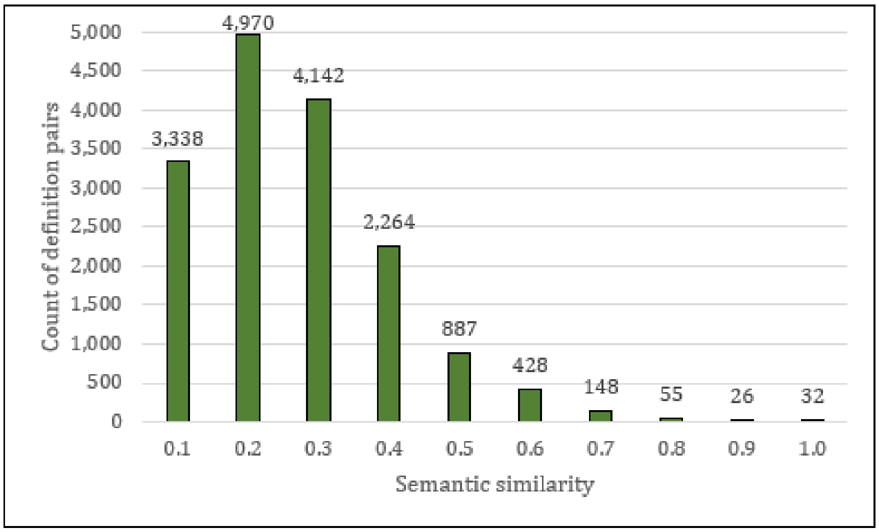

Figure 3: Histogram of the semantic similarity scores (range: 0 to 1, in which 1 is a maximum similarity) and counts of definition pairs. The pairwise semantics similarity scores among UE definitions are usually quite small.

similar divergence as the overall set of UE definitions, indicating that the inability to define UE is a fractal problem. This is supported by findings from prior research, including [37] noting that UE lacks an agreed-upon definition, Garett et al. [16] reviewing UE in website design, and Suh and Cheung [69] who reviewed 59 empirical studies of UE within various research communities.

From this foundation, we conducted a more focused analysis. Several categories of characteristics were identified among the 181 articles that explicitly defined UE. Table 5 presents this categorization of UE definitions, outlining their characteristics and providing examples. To establish these categories of UE and their characteristics, a two-step methodological approach was employed that drew from relevant techniques in thematic analysis. First, our *linguistic term analysis* [30] (described earlier in this section) provided a quantitative overview of the most frequently occurring terms; this served as a basis for identifying potential themes and categories related to UE. Building on the linguistic term analysis, a more *in-depth thematic analysis* was conducted to explore further and refine the emerging categories. This involved manually examining the definitions, contextual meanings, and relationships between the identified terms and themes. We actively discussed discrepancies from the linguistic term analysis to refine the categories as needed. This iterative process allowed for the development of a shared coding scheme that captured the nuances of UE across the definitions. Based on these considerations, three main Definition Categories (DCs) were ultimately established for further thematic coding:

DC1: User-System Interactivity (USI). User-system interactivity includes UB (towards the system), feedback, endurability, time, quality, user challenges, user control, and novelty.

DC2: User-System Perception (USP). User-system perception encompasses the user's cognitive and emotional responses, sensory

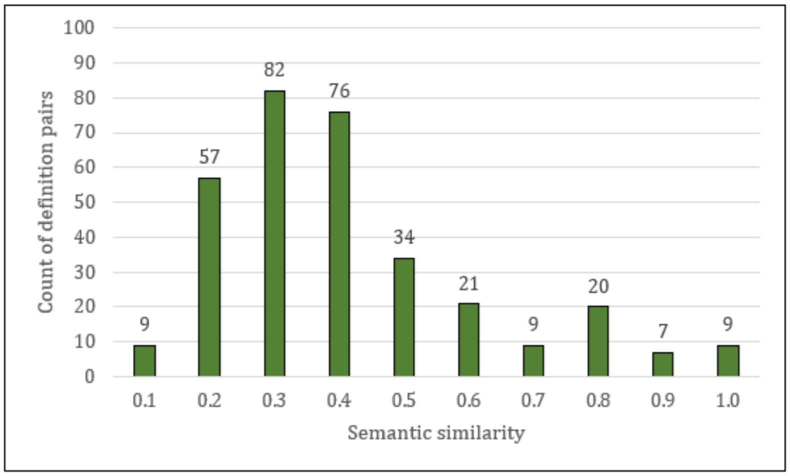

Figure 4: Histogram of the semantic similarity scores and counts of definition pairs by semantic similarity for the social media research. The pairwise semantics similarity scores among UE definitions are quite small in most cases, even within the same thematic area.

impressions, aesthetic appreciation, and overall value derived from the system.

DC3: User-Technology Affinity (UTA). User-technology affinity encompasses UB (towards the technology brand) and user loyalty, both as facets underlying the user's larger emotional, psychological, and physical investment in the system services provider.

Based on the coding results, Table 6 quantifies the variability of characteristics in definitions of UE. Definitions that did not fall into any categories and were research community-specific (i.e., articles that explicitly measured UE in alignment with specific goals of their thematic area, such as health) were also separately coded as such and further included in the table. The data comes from a subsample of 181 articles (out of n= 241 total; 75.1%) explicitly defining UE.

For **USI**, the most used definition characteristics were UB (system) in 114 articles (63.0%), followed by endurability in 38 articles (21.0%), and time, novelty, quality, feedback, user challenges, and user control in varying degrees. For **USP**, the affective characteristics were the most used in 103 articles (56.9%), followed by cognitive characteristics in 50 articles (27.6%), aesthetic characteristics in 17 articles (9.4%), sensory characteristics in 12 articles (6.6%), and general characteristics in 8 articles (4.4%). For **UTA**, UB (brand of technology) was the most used definition characteristic in 13 articles (7.2%), followed by user loyalty in 4 articles (2.2%).

The most included characteristic (63% of articles) fell under the category of USI, specifically UB, during interactions with the system. Affective characteristics of USP (e.g., "positive feelings" during user-system interactions) are also frequently emphasized (56.9% of articles), followed by cognitive (27.6%) and aesthetic aspects (9.4%) (e.g., users' attentiveness and appeal to system interface aesthetics, respectively), with sensory processes (e.g., touch) less commonly referenced (6.6%). Additionally, a substantial percentage of articles (34.8%) referred to thematic area-specific UBs and/or outcomes in their definitions of UE.

In the thematic-specific category, UB (within the thematic area, e.g., users' health behaviors in the health area) was the most used definition characteristic in 51 articles (28.2%), followed by user outcome (e.g., health outcomes) in 35 articles (19.3%). For example, studies describing UE in the context of mobile health applications included health-specific patient behaviors and/or clinical outcomes in their definitions of engagement.

Compared with other categories, we also observed that UTA is relatively underexplored or is an emerging area of research, with few articles considering UB toward brands (7.2%) or user loyalty (2.2%) as characteristics of UE. This may imply an emphasis on technical interactions in UE studies, i.e., UBs and perceptions are primarily viewed through the lens of direct user interactions with systems [27, 31] instead of their interactions with systems/service providers.

The mean number of characteristics included in definitions of UE is 3.2 characteristics per definition, with a median of 3.0, a maximum of 10.0, and a standard deviation of 2.0. These findings highlight the wide distribution of characteristics across UE definitions, reflecting the field's breadth and lack of consistency.

Together, these findings have important implications for understanding how UE is defined. Tables 2–6 highlight how UE has no consistent definition; instead, it is multifaceted and operationalized in various ways depending on the specific objectives of inquiry. The variability in defining UE leads to inconsistencies in measuring and interpreting what engagement is and the engagement levels. This lack of understanding of UE, including whether it even exists, hinders our understanding of its impact on UB and performance.

One possible explanation for the inconsistent nature of the UE definitions is that the concept itself is inherently flawed by attempting to capture a diverse range of behaviors, experiences, and interactions that may not share a common underlying mechanism or cause. In this sense, UE might be better understood as an *umbrella label* encompassing multiple, distinct phenomena rather than a singular, coherent construct. Such a perspective calls into question UE as a meaningful and measurable concept. Indeed, the present data suggest that UE researchers may need to reconsider how they

Table 5: Characteristics of UE definitions with examples.

Category	Category definitions	Characteristics	Characteristics definitions	Examples
User-System Interactivity (USI)	Exchange of inputs and outputs between user and system, respectively.	User Behavior (System)	Actions taken by the user to provide input to the system.	“[user’s] behavioural response that is prompted by the exposure” [63]
		Endurability	Repeated exchange of inputs and outputs between user and system, respectively.	“initial reactions to technologies and sustained use of and re-engagement with information systems over time” [51]
		Feedback	Output produced by the system in response to user input.	“the emotional, cognitive and behavioural connection that exists, at any point in time and possibly over time, between a user and a resource” [3]; “when they appreciate the (. . .) content to which they have given their attention” [74]; “quality of the user experience that emphasizes the positive aspects of the interaction” [1, 36–39, 42, 44, 46, 47, 47, 49, 60]; “quality of user experience characterized by attributes of challenge, positive affect, endurability, aesthetic and sensory appeal, attention, feedback, variety/novelty, interactivity, and perceived user control” (foundational definition by [54])
		Time	Duration of interactional exchanges (discrete or cumulative instances) between user and system.	
		Quality	System capacity to respond to user inputs consistently and successfully.	
		User Challenges	Obstacles to user-system interactivity, such as user navigation difficulties or system output failures.	
		User Control	Ability for users to navigate the system in alignment with user needs.	
		Novelty	System capacity to introduce variety to interactions.	
User-System Perception (USP)	User perceptions of the system.	Affective	User emotional processes (e.g., motivation, satisfaction) during interactions with the system.	
		Cognitive	User mental processes (e.g., attention) during interactions with the system.	
		Sensory	User sensory processes (e.g., tactile, auditory) during interactions with the system.	
		Aesthetic	User perception of the visual design of the system during interactions with the system.	
User-Technology Affinity (UTA)	Interactions and behavioral connections users have with technology platforms that incorporate a brand element, highlighting how brand perception of technology influences user-system engagement within HCI contexts	User Behavior (Brand)	Continuous actions taken by the service user to interact with a specific service provider’s systems and systems-related content (e.g., marketing materials), such as due to preference over other service providers.	“strong ties [with a brand]” [59]; “a behavioral manifestation that drives motivation to attach to the brand of the service provider beyond using the service” [29]
		User Loyalty		

approach the study of user-system interactions and experiences to be significantly more explicit to specific goals and measures.

The conflicting categories and characteristics identified in this research raise important doubts about the feasibility of developing a single, universally applicable measure of UE for HCI. Given the wide range of factors that contribute to UE and the thematic-specific nature of these factors, it is unlikely that a one-size-fits-all metric

Table 6: Characteristics of UE definitions. Coding was tallied based on the definitions' frequency of mutually exclusive occurrences of characteristics. As definitions could have multiple characteristics, articles can be in more than one category (N).

Definition characteristics (n= 181 articles)	N	% of N	% of 241 Articles
User-System Interactivity (USI)	N	% of N	% of 241 Articles
User Behavior (System)	114	49.1%	63.0%
Endurability	38	16.4%	21.0%
Time	16	6.9%	8.8%
Novelty	16	6.9%	8.8%
Quality	14	6.0%	7.7%
Feedback	12	5.2%	6.6%
User Challenges	11	4.7%	6.1%
User Control	9	3.9%	5.0%
General	2	0.9%	1.1%
Total	232	100.0%	
User-System Perception (USP)	N	% of N	% of Articles
Affective	103	54.2%	56.9%
Cognitive	50	26.3%	27.6%
Aesthetic	17	8.9%	9.4%
Sensory	12	6.3%	6.6%
General	8	4.2%	4.4%
Total	190	100.0%	
User-Technology Affinity (UTA)	N	% of N	% of Articles
User Behavior (Brand of Technology)	13	76.5%	7.2%
User Loyalty	4	23.5%	2.2%
Total	17	100.0%	
Thematic-specific	N	% of N	% of Articles
User Behavior (Research Focus)	51	59.3%	28.2%
User Outcome	35	40.7%	19.3%
Total	86	100.0%	

could accurately capture the complexity of UE across different thematic areas and settings. This brings us to our second RQ.

4.3 RQ2: How Is User Engagement Measured?

Our analysis of empirical UE metrics further confirmed our concerns with various UE definitions within HCI research. This subdata was identified from 222 articles (of n = 241 total sample) that provided a metric of UE (see Table 7).

Results in Table 7 show how UE is most commonly measured through self-reported and activity-based metrics. Specifically, self-reported metrics, such as questionnaires or surveys, were used in 34.4% (n= 83) of the articles, while activity-based metrics, such as time spent on a website or the number of posts made (e.g., on an online social network) were leveraged in more than half (n= 125, 51.9%) of the articles. System analytics, such as total user base or daily user volume, were adopted in only 10.4% (n=25) of the articles. So, some researchers use more objective metrics, while others use more subjective metrics to measure UE.

There were also combinations of different metric types used in the articles. Sixteen articles (6.6%) used both self-reported and activity-based metrics, one paper (0.4%) used self-reported and system analytics metrics, and 14 articles (5.8%) used activity-based and system analytics metrics. A small percentage (1.2%) of articles used a combination of all three categories of metrics (self-report, activity-based, and system analytics). So, in total, 34 out of 241 articles (14.1%) used different metric types (i.e., multiple ways of measuring engagement), whereas the majority (n= 188, 78.0%) subscribed to only one way of measuring UE. This finding suggests that researchers use one and occasionally two metrics (e.g., combined self-report and activity-based metrics, 6.6%) instead of a comprehensive approach. The reliance solely on self-report (34.0%) or activity-based (38.2%) metrics highlights the need for researchers to develop more objective and standardized measures of UE. Also, the employment of activity-based metrics (i.e., UBs and actions within a product, service, or system), used in 38.2% of UE studies, highlights that 'system use' is the most important set of metrics.

As we surmised earlier, the lack of a consistent and universally accepted definition of UE is likely to create confusion and misinterpretation among HCI researchers and practitioners. As we observe here, measures are often taken as some form of proxy that

Table 7: Empirical metrics of UE. The frequency of mutually exclusive metrics was coded by tallying for frequency in each paper that measured UE.

Empirical metrics (222 articles)	N=222 Articles	% of N	% of 241 Articles
Activity-based	125	53.6%	51.9%
Self-reported	83	35.6%	34.4%
System Analytics	25	10.7%	10.4%
Total	233	100.0%	100.0%
Metrics combinations (222 articles)	N=222 Articles	% of N	% of 241 Articles
Activity-based only	92	41.4%	38.2%
Self-reported only	82	36.9%	34.0%
Self-reported and Activity-based	16	7.2%	6.6%
Activity-based and System Analytics	14	6.3%	5.8%
System Analytics only	7	3.2%	2.9%
Self-reported, Activity-based, and System Analytics	3	1.4%	1.2%
Self-reported and System Analytics	1	0.5%	0.4%
N/A	7	3.2%	2.9%
Total	222	100.0%	100.0%

Notes: Activity-based = Tracking UBs and actions within a product, service, or system, such as the number of social media posts on a website.
Self-reported = Individuals answer questions about their level of engagement with a product, service, or system, such as via surveys.
System Analytics = Data collected from the system or platform, such as daily user volume.

researchers then often loosely relate to UE. The ambiguity in defining UE leads to the use of many proxies, making it difficult to draw meaningful conclusions about the impact or relevance of UE on the studies. In particular, diverse and sometimes contradictory metrics at the aggregate level suggest that researchers may measure different aspects of UB and experiences rather than a singular, cohesive concept of UE. This fragmentation of UE measures hinders the ability to compare and synthesize findings across studies and capture both specific and/or spectrums of psychological aspects of technology use.

One potential explanation for the myriad measures of UE that exist is the inherently context-dependent nature of the concept. UE may manifest differently across various platforms, industries, and user populations, making it challenging to identify a universally applicable definition or set of metrics. This context-dependent nature of UE further complicates the development of standardized measurement approaches, as researchers must account for the unique features and dynamics of each specific context in which UE is being studied. Indeed, the significant reliance on self-reported and activity-based metrics implies difficulty in drawing meaningful conclusions from UE research. Furthermore, as our thematic-specific analysis of social media shows, even within a research topic, UE is loosely defined.

The discrepancy in the measurement approaches goes back to the arbitrary, even paradoxical nature of UE. For example, consider users who click a lot. Do the users click often because they are "engaged" positively, enjoying the experience, and finding the system helpful? Or do users click often because they cannot locate information or understand how the system works? A solely objective measure of clicks would give an incomplete picture of UE without the necessary side information to contextualize it. The different ways of measuring UE indicate no consensus on the answer to this paradox. Furthermore, the relatively small share of studies applying a combination of different measurement types indicates that researchers struggle to commensurate these two views—behavioral or latent construct—into a coherent whole. The affective or emotional component of the UE definition (as well as the cognitive aspect) would suggest that researchers need to inquire about latent aspects of UE to understand better its formation; otherwise put, we could measure the activities a user takes with a system by counting their frequency, but we are unable to infer from the count data whether this use was pleasurable or not, whether the user was focused or clicked randomly, and so on. So, a major gap is a lack of unification in terms of behavioral and latent views of UE as currently put forth in the literature.

4.4 RQ3: Are Conceptualization and Assessments of User Engagement Aligned Across Studies?

Results in Table 8 demonstrate the significant heterogeneity in how UE is defined by summarizing our dataset's different combinations of characteristics used across UE definitions.

One of the most striking patterns is the prevalence of definitions incorporating conceptualizations from USI and USP categories, such as UB (towards the system) and affective perceptions (see Table 8). Table 8 also includes definitions referencing other conceptualizations from these larger categories, such as endurability, cognitive, and sensory components. It is also worth highlighting how a significant portion of studies solely consider USP characteristics (e.g., affective only, 9.4%, or both affective and cognitive, 3.9%), indicating

Table 8: Combinations of conceptualization of UE definitions. Note: The table includes data from a subset of studies (n= 147). Combinations of attributes in the remaining studies with definitions only occurred once (i.e., each provided unique definitions, such as thematic-specific terminology) and were thus excluded from the table. The '+' symbol in Table 8 indicates combining the conceptualizations).

No.	Combination of Conceptualizations (Across Categories)	Definition Category			Frequency of articles	%
		User-System Interactivity	User-System Perception	User-Technology Affinity		
1	User Behavior (System)	✓			38	21.0
2	User Behavior (System) + Affective	✓	✓		22	12.2
3	Affective only		✓		17	9.4
4	User Behavior (System) + Endurability + Affective + Cognitive	✓	✓		9	5.0
5	Affective + Cognitive		✓		7	3.9
6	Thematic-specific only (no category)		N/A		7	3.9
7	User Behavior (System) + Affective + Cognitive	✓	✓		6	3.3
8	User Behavior (System) + Endurability	✓			6	3.3
9	Affective + User Behavior (Brand)		✓	✓	5	2.8
10	User Behavior (System) + Feedback + Time + User Challenges + User Control + General + Affective + Cognitive + Sensory + Aesthetic	✓	✓		4	2.2
11	User Behavior (System) + Time + Affective + Cognitive	✓	✓		4	2.2
12	Endurability only	✓			4	2.2
13	User Behavior (Brand) only			✓	3	1.7
14	User Behavior (System) + Feedback + Endurability + User Challenges + User Control + Novelty + Affective + Cognitive + Sensory + Aesthetic	✓	✓		2	1.1
15	User Behavior (System) + Endurability + Time + Affective + Cognitive	✓	✓		2	1.1
16	User Behavior (System) + Time + Affective + Cognitive	✓	✓		2	1.1
17	User Behavior (System) + Time + Affective	✓	✓		2	1.1
18	User Behavior (System) + Affective + General	✓	✓		2	1.1
19	User Behavior (System) + Endurability + Affective	✓	✓		2	1.1
20	Endurability + Affective	✓	✓		2	1.1
21	Endurability + Time	✓			2	1.1
22	Sensory only		✓		2	1.1
23	General only		✓		2	1.1

that UE is also sometimes considered in terms of subjective assessments without any direct reference to user-system (or user-brand) interactivity.

In sum, we identify 23 different characteristic combinations for UE, with the most prevalent being User Behavior (System), with 21.0% being sole 'use of a system.' For the other 79.0% of UE approaches, there were 22 measurement combinations. These findings clearly show that UE has no clear operationalization in the literature, and it is *de facto* being used as a catch-all for whatever system usage metrics researchers deem to measure. The results highlight considerable variability in how UE is defined and measured, reinforcing the need for a more cohesive approach within the HCI field. In the next section, we discuss these findings in the context of their implications for both researchers and practitioners, emphasizing the contributions of this study to the ongoing discourse of UE with HCI research.

5 Discussion

5.1 Discussion of the Findings

Returning the findings, beginning with how the literature defines UE (RQ1). UE is mainly defined as the quality of UX and emotional, cognitive, and behavioral connection at any point in time between a user and a resource, usually a piece of technology [see, for example, 39]. Past HCI researchers have highlighted a continual call for more standard operationalization of UE [52]. Specifically, researchers noted the need for standardization and consensus in defining UE to facilitate consistent measurement and analysis across different research communities and contexts [see, for example, 37].

Since a consistent universal UE definition does not exist, its measurement (RQ2) tends to be arbitrary, roughly divisible into behavioral (direct) and self-reported (indirect) measurements. However, neither approach can fully satisfy the given definitions in the field. Therefore, UE's value as a psychological construct is limited. While abstract concepts such as love, trust, and satisfaction are valuable in social sciences, they benefit from reasonably clear operational definitions and consistent measurement approaches. In contrast, UE has *yet to achieve* this level of theoretical clarity, hindering its applicability as a practical concept for system design and behavioral user research.

As such, a key epistemic challenge of current UE research is the issue of construct validity. A lack of a clear and universally accepted definition of UE has led to inconsistency in how it is operationalized and measured. This, in turn, raises questions about the *validity* and *reliability* of UE as a construct. If the definitions and measurements used to study UE vary significantly across studies, it becomes difficult to determine whether the results obtained indeed capture the intended construct. Consequently, this poses challenges in accumulating knowledge about UE and limits the development of a coherent body of literature on the subject.

Due to inconsistent conceptualizations of UE across studies, researchers have called for more comprehensive and standardized metrics that capture both objective and subjective aspects of UE. However, self-reported metrics can be biased [6], while activity-based metrics may not capture all aspects of what is considered UE. Meanwhile, system analytics [41] can provide objective data, but their limitations critically include the inability to capture subjective user perceptions. Research has attempted to develop new methods to measure UE through self-reported, activity-based, and system analytics metrics. Such an approach has attempted to enable more holistic yet precise UE evaluations and facilitate critical metrics validation across different user-system interactions. The result has been a complex combination of affective, cognitive, and behavioral metrics covering nearly all aspects of users' interactions with technology, thus not bringing clarity but returning to the "source" of the UE concept, which is *using a system*.

Moving to RQ3, this heterogeneity again raises questions about the practical applicability of the concept of UE. As current definitions of UE inconsistently reference various technical and subjective dimensions of user activity, measuring engagement accurately or applying findings to other contexts remains challenging. Without consensus on what specific characteristics should be included across contexts (whether based on metrics of relevance or significance), UE studies often remain too broad and multifaceted to be meaningfully compared. Researchers may also inadvertently study different aspects of UB and UX [40], leading to inconsistent findings and an inability to synthesize results across studies. This ambiguity hinders the scientific advancement of UE research and the science surrounding it and limits its practical application in various HCI research communities. Notably, the epistemic challenges of UE research extend to the issue of replicability. The inconsistencies in defining and measuring UE hinder the ability of HCI researchers to reproduce the findings of previous studies accurately, as it is difficult to determine whether researchers are truly replicating the methods used in the original study and, thus, whether any discrepancies in the results can be attributed to methodological differences or genuine inconsistencies in the underlying UE phenomena.

In short, our findings highlight the significant challenges researchers and practitioners face due to the lack of a consistent UE definition, the diverse range of empirical metrics, and the conflicting categories and characteristics of UE—all of which contribute to a fragmented and disorganized understanding of UE. Our analysis suggests that UE has no accepted clear definition, even after decades of HCI research. With no clear definition, UE struggles to be clearly measured, and perhaps more importantly, the concept does not live up to its full potential of helping us understand how and why users use a system. With ambiguous definitions and metrics, UE has been measured via various contrasting, inconsistent, and sometimes conflicting dimensions. However, in the literature, UE is often presented as well-definable, readily measurable, and explicitly supportive of user understanding and HCI. There is an analogy here to many abstract constructs in social sciences, such as trust, love, and satisfaction, that also lack tangible manifestations yet provide significant value in understanding human behavior. However, unlike these constructs, which benefit from established definitions and theoretical foundations, UE remains an umbrella term with conflicting definitions and no clear consensus on measuring or applying it. While our argument does not deny the potential of UE as an abstract construct, we emphasize the need for operationalization to make it a meaningful tool for scientific and practical development in the HCI research communities.

User engagement is a set of measures defined by the researcher within the specific context of studying the use of a system by one or more people.

5.2 Implications for Researchers and the HCI Scientific Community

Consistent conceptualizations of UE within HCI would benefit the field by enhancing comparability and synthesis of research findings across HCI studies or at least within studies within research communities. This study focused on UE as it applies specifically to user interactions with technology, aligning with general areas of HCI research. The development of HCI-centric definitions of and metrics for UE would allow researchers within research communities (i.e., information retrieval, virtual reality, health, social media, etc.) to create frameworks that address user-system interactions' unique characteristics and applications. Given that HCI is a broad and multidisciplinary field, this targeted approach for a consistent UE concept leverages its inherent strengths, supports diverse methodologies, and improves the impact of HCI research communities.

We provide four (what we believe are) helpful guidelines (GL) for HCI researchers on how to think about UE.

GL01: When comparing your results to other UE studies, be cautious about drawing direct comparisons if the UE operational definition or metrics are substantially different. This helps avoid overgeneralizing findings based on narrow operational definitions and inconsistent measurement. HCI researchers should assess when UE metrics are commensurable and when they are not. This also involves clearly defining and operationalizing the specific UE metrics you are measuring in your study to facilitate cross-study comparisons.

GL02: Focus on developing and validating more precise, context-specific measures for UE in your thematic or application area. Break UE into specific, measurable components such as attention, cognitive processing, or behavioral indicators relevant to your research context. When reporting results, specify what was measured and how it relates to concrete aspects of UB or UX. Such divergent operationalization of the UE construct can benefit incremental knowledge accumulation, as it eradicates the need for research, e.g., to "reinvent the wheel" of what UE means and how it is measured. This approach helps articulate scientific progress in studies about UE, akin to many computer science research communities using baselines, benchmarks, and commonly agreed metrics for measurement of progress in computational tasks [73]. This effort should be coupled with documenting contextual factors affecting UE to support incremental knowledge accumulation.

GL03: Employ UE as a *categorical* label. UE has value as a categorical label (a.k.a. umbrella construct) across HCI communities for a range of user-system measures, metrics, and concepts as defined within the context of a given HCI study. In this context, UE has several practical implications. The phrase "user engagement" might help ease communication within a research article or study. As such, the categorical label version of UE can be easily defined.

GL04: Community operationalization of UE. HCI research communities using UE would benefit from standardized core practices. To accomplish this, there is a host of actions inspired by discourse in other scientific domains and, more broadly, in the philosophy of science [34] that HCI communities engaged in UE research could draw from, including common (1) common operational definitions, (2) standardized measurement and protocols, (3) reporting guidelines for UE metrics, (4) databases of UE measurements, and (5) shared exemplars and method. This community-specific discussion and derivative social practices are mainly missing from how HCI research communities currently approach UE. As usability can be viewed as only existing within given contexts [8], UE could also benefit from this contextualize within communities. For example, defining and measuring UE might differ between virtual reality platforms and websites. Social practices within the HCI communities (or lack thereof) play a role in UE standardization because these practices are critical interactions in scientific progress [43]. So, we believe this community-specific perspective on UE standardization would be more fruitful than a forced *universal* attempt at a broad UE standardization across all HCI.

5.3 Strengths, Limitations, and Future Work

The strengths of the current research lie in its comprehensive analysis of UE definitions and metrics in the HCI literature, which is anchored in the rigor employed in the systematic literature search and the combination of linguistic and thematic analyses. By highlighting the inconsistencies and lack of standardized measures for UE, this study also raises important questions about the scientific and practical applicability of the UE concept within HCI research.

Notwithstanding, there are also limitations to this current work. First, while the systematic review captures a broad range of HCI studies, it may not encompass all relevant work due to excluding research like customer studies that may touch on user aspects. Future research should explore the potential for refining UE into a more coherent construct within HCI communities. Future HCI studies should investigate alternative UE metrics that can provide more precise and actionable insights beyond the current ambiguous application of UE. Also, the linguistic and thematic analyses are inherently interpretive; other researchers may identify different patterns or themes within the literature. We did take steps to mitigate this concern by assessing interrater reliability and ensuring high consistency between independent coders. Relatedly, we have not considered the design of systems [26], which views user involvement in participation design. We have also not considered adjacent domains, such as business [7, 24], which is attentive to a company's efforts focused on loyalty and awareness.

There is much future work that HCI researchers, most notably those who may wish to pursue UE research, can explore. First, investigating the 'engagement' aspect of UE can be clarified by borrowing from concepts such as cognitive absorption, focused immersion, focused concentration, flow, or mindfulness. UE could be compared with these conceptualizations of engagement. Perhaps engagement in these areas of study has been better conceptualized. However, O'Brien and Toms [54] attempted this using related theories of UB and psychology, like flow, play, aesthetics, and information interaction, as did Peters et al. [57]. These authors examine different definitions and perspectives on engagement from various domains and categorize them according to the perception-cognition-action loop. Based on the current analysis, significantly more work in this area is still needed.

Second, a detailed analysis of the evolution of UE definition over the years through the lens of the different contexts or platforms could be carried out. Perhaps UE is more clearly defined within a narrow context. While possibly fruitful, however, our similarity analysis within social media shows this is not the case. UE is still broadly defined within this narrower area, which supports our HCI community-level view of UE development. So, this is an avenue for much future research.

Third, as our research highlights the divergence, perhaps a fruitful area of HCI research could be examples or case studies where UE has been measured consistently within the same context. If these exist, the findings will help understand if there is inconsistency within similar contexts. Based on our findings, this is not the case. Fourth, identifying patterns in how UE is measured in specific contexts or on certain platforms and whether this measurement has been effective. Such findings could help standardize UE measurement for particular research contexts. However, our findings indicate that there is no such standardization as of present. Finally, the concept of UE, as existing in the prior literature, needs investigating with the introduction of large language models that empower technology with dialogue and conversational capabilities [4, 86]. In summary, future research should aim to move UE from construct to label and develop community-specific UE definitions that provide more explicit guidance for academic inquiry and practical application with HCI research communities. These approaches would help mitigate confusion and enhance the utility of UE in HCI. In the near term, UE as a categorical label seems the most fruitful overall, while HCI research communities develop divergent views and community-specific implementations of UE.

6 Conclusion

This research examines how UE has been employed in prior HCI work and recommends UE as a label for a set of study-specific metrics to understand UB within HCI research better. Acknowledging UE as a categorical label could allow for a more flexible and context-specific approach to studying the use of a system. This shift in perspective enables a more adaptable approach to studying UE and encourages HCI researchers to identify and investigate research community-specific UE and system design. By doing so, HCI researchers can draw more meaningful insights, develop targeted recommendations for system improvements, enhance research across various HCI communities, and then work to operationalize thematic-specific UE implementation toward higher relevance and applicability. The plurality observed in our study concerning the UE construct supports a community-driven approach to UE. HCI research communities should focus effort and attention on formalizing what UE means for them, how it is measured, and what the shared norms, practices, and resources are for UE operationalization.

References

[1] Maristella Agosti, Nicola Orio, and Chiara Ponchia. 2018. Promoting user engagement with digital cultural heritage collections. *International Journal on Digital Libraries 19*, 353–366. https://doi.org/10.1007/s00799-018-0245-y

[2] Aamir Amin, Shuib Basri, Mohd Fadzil Hassan, and Mobashar Rehman. 2018. A Snapshot of 26 Years of Research on Creativity in Software Engineering - A Systematic Literature Review. In *Mobile and Wireless Technologies 2017* (*Lecture Notes in Electrical Engineering*), 2018. Springer, Singapore, 430–438. https://doi.org/10.1007/978-981-10-5281-1_47

[3] Simon Attfield, Gabriella Kazai, Mounia Lalmas, and Benjamin Piwowarski. 2011. Towards a science of user engagement (position paper). In *WSDM Workshop on User Modelling for Web Applications*, February 09, 2011. Hong Kong, China.

[4] Carina Benz, Lara Riefle, and Gerhard Satzger. 2024. User Engagement and Beyond: A Conceptual Framework for Engagement in Information Systems Research. *Communications of the Association for Information Systems* 54, (January 2024), 331–359. https://doi.org/10.17705/1CAIS.05412

[5] Kerstin Bongard-Blanchy, Arianna Rossi, Salvador Rivas, Sophie Doublet, Vincent Koenig, and Gabriele Lenzini. 2021. " I am Definitely Manipulated, Even When I am Aware of it. It's Ridiculous!"-Dark Patterns from the End-User Perspective. In *Designing Interactive Systems Conference 2021*, 2021. 763–776.

[6] John Bound, Charles Brown, and Nancy Mathiowetz. 2001. Chapter 59 - Measurement Error in Survey Data. In *Handbook of Econometrics*, James J. Heckman and Edward Leamer (eds.). Elsevier, 3705–3843. https://doi.org/10.1016/S1573-4412(01)05012-7

[7] Roderick J. Brodie, Linda D. Hollebeek, Biljana Jurić, and Ana Ilić. 2011. Customer engagement: Conceptual domain, fundamental propositions, and implications for research. *Journal of Service Research* 14, 3 (2011), 252–271.

[8] john Brooke. 1996. SUS: A "Quick and Dirty" Usability Scale. In *Usability Evaluation In Industry*. CRC Press.

[9] Daniel Cer, Yinfei Yang, Sheng-yi Kong, Nan Hua, Nicole Limtiaco, Rhomni St John, Noah Constant, Mario Guajardo-Cespedes, Steve Yuan, Chris Tar, Yun-Hsuan Sung, Brian Strope, and Ray Kurzweil. 2018. Universal Sentence Encoder. https://doi.org/10.48550/arXiv.1803.11175

[10] Chris Chapman, Edwin Love, Russell P. Milham, Paul ElRif, and James L. Alford. 2008. Quantitative Evaluation of Personas as Information. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, September 01, 2008. 1107–1111. https://doi.org/10.1177/154193120805201602

[11] Chris Chapman and Russell P. Milham. 2006. The Personas' New Clothes: Methodological and Practical Arguments against a Popular Method. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, October 01, 2006. 634–636. https://doi.org/10.1177/154193120605000503

[12] Neha Chaudhuri, Gaurav Gupta, and Aleš Popovič. 2025. Do you believe it? Examining user engagement with fake news on social media platforms. *Technological Forecasting and Social Change* 212, (2025), 123950.

[13] Sara J. Czaja and Joseph Sharit. 2003. Practically relevant research: Capturing real world tasks, environments, and outcomes. *The Gerontologist* 43, suppl_1 (2003), 9–18.

[14] Kevin Doherty and Gavin Doherty. 2019. Engagement in HCI: Conception, Theory and Measurement. *ACM Comput. Surv.* 51, 5 (September 2019), 1–39. https://doi.org/10.1145/3234149

[15] Jeffrey R. Edwards and Richard P. Bagozzi. 2000. On the nature and direction of relationships between constructs and measures. *Psychological Methods* 5, 2 (2000), 155.

[16] Renee Garett, Jason Chiu, Ly Zhang, and Sean D. Young. 2016. A literature review: website design and user engagement. *Online Journal of Communication and Media Technologies* 6, 3 (2016), 1.

[17] Barney G. Glaser and Anselm L. Strauss. 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Transaction Publishers.

[18] Anne-Wil Harzing. 2010. *The Publish or Perish Book*. Tarma Software Research Pty Limited Melbourne, Australia. Retrieved September 7, 2024 from http://www.harzing.info/download/popbook12.pdf

[19] Paul M. Hirsch and Daniel Z. Levin. 1999. Umbrella Advocates Versus Validity Police: A Life-Cycle Model. *Organization Science* 10, 2 (April 1999), 199–212. https://doi.org/10.1287/orsc.10.2.199

[20] Harry Hochheiser and Jonathan Lazar. 2007. HCI and societal issues: A framework for engagement. *International Journal of Human Computer Interaction* 23, 3 (2007), 339–374.

[21] Bernhard Hommel, Craig S. Chapman, Paul Cisek, Heather F. Neyedli, Joo-Hyun Song, and Timothy N. Welsh. 2019. No one knows what attention is. *Atten Percept Psychophys* 81, 7 (October 2019), 2288–2303. https://doi.org/10.3758/s13414-019-01846-w

[22] Kasper Hornbæk and Antti Oulasvirta. 2017. What is interaction? In *Proceedings of the 2017 CHI conference on human factors in computing systems*, 2017. 5040–5052. Retrieved December 8, 2024 from https://dl.acm.org/doi/abs/10.1145/3025453.3025765

[23] Daniel J. Hruschka, Deborah Schwartz, Daphne Cobb St John, Erin Picone-Decaro, Richard A. Jenkins, and James W. Carey. 2004. Reliability in coding open-ended data: Lessons learned from HIV behavioral research. *Field Methods* 16, 3 (2004), 307–331.

[24] Chih-Lung Hung, Jen-Her Wu, Pei-Yu Chen, Xiaoyu Xu, Wan-Ling Hsu, Li-Min Lin, and Ming-Che Hsieh. 2023. Enhancing healthcare services and brand engagement through social media marketing: Integration of Kotler's 5A framework with IDEA process. *Information Processing & Management* 60, 4 (July 2023), 103379. https://doi.org/10.1016/j.ipm.2023.103379

[25] Jörn Hurtienne. 2017. How Cognitive Linguistics Inspires HCI: Image Schemas and Image-Schematic Metaphors. *International Journal of Human–Computer Interaction* 33, 1 (January 2017), 1–20. https://doi.org/10.1080/10447318.2016.1232227

[26] Mark I. Hwang and Ron G. Thorn. 1999. The effect of user engagement on system success: a meta-analytical integration of research findings. *Information & Management* 35, 4 (1999), 229–236.

[27] Bernard J. Jansen, Mimi Zhang, and Carsten D. Schultz. 2009. Brand and its effect on user perception of search engine performance. *J. Am. Soc. Inf. Sci.* 60, 8 (August 2009), 1572–1595. https://doi.org/10.1002/asi.21081

[28] Tingting Jiang. 2014. Exploratory Search: A Critical Analysis of the Theoretical Foundations, System Features, and Research Trends. In *Library and Information Sciences: Trends and Research*, Chuanfu Chen and Ronald Larsen (eds.). Springer, Berlin, Heidelberg, 79–103. https://doi.org/10.1007/978-3-642-54812-3_7

[29] Soohyung Joo, Namjoo Choi, and Tae Hyun Baek. 2018. Library marketing via social media: The relationships between Facebook content and user engagement in public libraries. *Online Information Review* 42, 6 (2018), 940–955. https://doi.org/10.1108/OIR-10-2017-0288

[30] S. Jothilakshmi and V. N. Gudivada. 2016. Chapter 10 - Large Scale Data Enabled Evolution of Spoken Language Research and Applications. In *Handbook of Statistics*, Venkat N. Gudivada, Vijay V. Raghavan, Venu Govindaraju and C. R. Rao (eds.). Elsevier, 301–340. https://doi.org/10.1016/bs.host.2016.07.005

[31] Nayeon Kim. 2024. Capturing Initial Gaze Attraction in Branded Spaces Through VR Eye-Tracking Technology. *International Journal of Human–Computer Interaction* (May 2024), 1–14. https://doi.org/10.1080/10447318.2024.2351717

[32] Eric I. Knudsen. 2007. Fundamental Components of Attention. *Annu. Rev. Neurosci.* 30, 1 (July 2007), 57–78. https://doi.org/10.1146/annurev.neuro.30.051606.094256

[33] Theodora Koulouri, Stanislao Lauria, and Robert D. Macredie. 2016. Do (and Say) as I Say: Linguistic Adaptation in Human–Computer Dialogs. *Human–Computer Interaction* 31, 1 (January 2016), 59–95. https://doi.org/10.1080/07370024.2014.934180

[34] Thomas S. Kuhn. 2012. *The Structure of Scientific Revolutions: 50th Anniversary Edition*. University of Chicago Press, Chicago, IL. Retrieved November 21, 2024 from https://press.uchicago.edu/ucp/books/book/chicago/S/bo13179781.html

[35] Neha Kumar and Nicola Dell. 2018. Towards Informed Practice in HCI for Development. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW (November 2018), 1–20. https://doi.org/10.1145/3274368

[36] Vesna Kumbaroska and Pece Mitrevski. 2017. Exploring User Engagement in Information Networks: Behavioural–based Navigation Modelling, Ideas and Directions. *BRAIN. Broad Research in Artificial Intelligence and Neuroscience* 7, 4 (2017), 93–107.

[37] Mounia Lalmas, Heather O'Brien, and Elad Yom-Tov. 2015. *Measuring User Engagement*. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-031-02289-0

[38] Janette Lehmann, Mounia Lalmas, Ricardo Baeza-Yates, and Elad Yom-Tov. 2013. Networked user engagement. In *Proceedings of the 1st Workshop on User Engagement Optimization*, November 2013. ACM, San Francisco California USA, 7–10. https://doi.org/10.1145/2512875.2512877

[39] Janette Lehmann, Mounia Lalmas, Elad Yom-Tov, and Georges Dupret. 2012. Models of User Engagement. In *User Modeling, Adaptation, and Personalization*, Judith Masthoff, Bamshad Mobasher, Michel C. Desmarais and Roger Nkambou (eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 164–175. https://doi.org/10.1007/978-3-642-31454-4_14

[40] Yuelin Li, Pengfei He, Xiaojun Yuan, and Frank Hopfgartner. 2024. Developing and evaluating a gamified information retrieval system for Generation Z. *Library & Information Science Research* 46, 1 (January 2024), 101270. https://doi.org/10.1016/j.lisr.2023.101270

[41] Yuelin Li and Chang Liu. 2019. Information Resource, Interface, and Tasks as User Interaction Components for Digital Library Evaluation. *Information Processing & Management* 56, 3 (May 2019), 704–720. https://doi.org/10.1016/j.ipm.2018.10.012

[42] L Liang. 2017. Essays on user acquisition, retention, experience, engagement and purchase decisions of a hybrid mobile wallet. University of Texas, University of Texas. Retrieved from https://repositories.lib.utexas.edu/handle/2152/63040

[43] Helen E. Longino. 1997. Interpretation versus Explanation in the Critique of Science. *Science in Context* 10, 1 (1997), 113–128.

[44] Juan Carlos Medina Serrano, Orestis Papakyriakopoulos, and Simon Hegelich. 2020. Exploring political ad libraries for online advertising transparency: lessons from Germany and the 2019 European elections. In *SMSociety'20: International Conference on Social Media and Society*, 2020. 111–121.

[45] Olimpia Meglio and Svante Schriber. 2020. Umbrella Constructs: An Overarching Framework. In *Mergers and Acquisitions*. Springer International Publishing, Cham, 27–56. https://doi.org/10.1007/978-3-030-40459-8_2

[46] K Nelissen. 2019. Essays on measuring user engagement with online news. Volg KU Leuven, Volg KU Leuven. Retrieved from https://lirias.kuleuven.be/2842743?limo$=$0

[47] Klaas Nelissen, Monique Snoeck, Seppe Vanden Broucke, and Bart Baesens. 2018. Swipe and Tell: Using Implicit Feedback to Predict User Engagement on Tablets. *ACM Transaction on Information Systems* 36, 4 (June 2018). https://doi.org/10.1145/3185153

[48] Michelle M. Ng, Joseph Firth, Mia Minen, and John Torous. 2019. User engagement in mental health apps: a review of measurement, reporting, and validity. *Psychiatric Services* 70, 7 (2019), 538–544.

[49] T Nguyen and T Le. 2013. *User Base Growth via Viral Marketing: The Contributory Factors: case: Ostosnero by Avansera Ltd.* theseus.fi. Retrieved from https://www.theseus.fi/handle/10024/57812

[50] Heather O'Brien. 2016. Theoretical perspectives on user engagement. In *Why engagement matters: Cross-disciplinary perspectives of user engagement in digital media*. Spring Nature, 1–26.

[51] Heather O'Brien and Paul Cairns. 2015. An empirical evaluation of the User Engagement Scale (UES) in online news environments. *Information Processing & Management* 51, 4 (2015), 413–427. https://doi.org/10.1016/j.ipm.2015.03.003

[52] Heather L. O'Brien, Paul Cairns, and Mark Hall. 2018. A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. *International Journal of Human-Computer Studies* 112, (2018), 28–39. https://doi.org/10.1016/j.ijhcs.2018.01.004

[53] Heather L. O'Brien, Ido Roll, Andrea Kampen, and Nilou Davoudi. 2022. Rethinking (Dis)engagement in human-computer interaction. *Computers in Human Behavior* 128, (March 2022), 107109. https://doi.org/10.1016/j.chb.2021.107109

[54] Heather L. O'Brien and Elaine G. Toms. 2008. What is user engagement? A conceptual framework for defining user engagement with technology. *Journal of the American Society for Information Science and Technology* 59, 6 (2008), 938–955. https://doi.org/10.1002/asi.20801

[55] Heather L. O'Brien and Elaine G. Toms. 2010. The Development and Evaluation of a Survey to Measure User Engagement. *Journal of the American Society for Information Science and Technology* 61, 1 (January 2010), 50–69. https://doi.org/10.1002/asi.21229

[56] J Oh, S Sudarshan, JA Lee, and N Yu. 2021. Serendipity enhances user engagement and sociality perception: the combinatory effect of serendipitous movie suggestions and user motivations. *Behaviour & Information* 41, 11 (2021), 2324–2341. https://doi.org/10.1080/0144929X.2021.1921027

[57] Christopher Peters, Ginevra Castellano, and Sara de Freitas. 2009. An Exploration of User Engagement in HCI. In *Proceedings of the International Workshop on Affective-Aware Virtual Agents and Social Robots* (*AFFINE '09*), 2009. New York, NY, USA. https://doi.org/10.1145/1655260.1655269

[58] Dorian Peters, Rafael A. Calvo, and Richard M. Ryan. 2018. Designing for motivation, engagement and wellbeing in digital experience. *Frontiers in Psychology* 9, (2018), 797.

[59] RE Puspita and S Saifudin. 2020. Creating Sharia Capital Market Literation Strategy on Social Media among Young Moslem. *Journal of Islamic Economics Lariba* (2020). Retrieved from https://journal.uii.ac.id/JIELariba/article/view/15112

[60] Sarifah Putri Raflesia, Kridanto Surendro, and Rossi Passarella. 2017. The user engagement impact along information technology of infrastructure library (ITIL) adoption. In *2017 International Conference on Electrical Engineering and Computer Science (ICECOS)*, 2017. IEEE, 184–187.

[61] William C. Robertson. 2002. *Force and Motion: Stop Faking It! Finally Understanding Science So You Can Teach It*. NSTA Press.

[62] Joni Salminen, Soon-gyo Jung, Ahmed Kamel, Willemien Froneman, and Bernard J. Jansen. 2022. Who is in the sample? An analysis of real and surrogate users as participants in user study research in the information technology fields. *PeerJ Comput. Sci.* 8, (October 2022), e1136. https://doi.org/10.7717/peerj-cs.1136

[63] Petra N. C. Sanne and Melanie Wiese. 2018. The theory of planned behaviour and user engagement applied to Facebook advertising. *South African Journal of Information Management* 20, 1 (June 2018). https://doi.org/10.4102/sajim.v20i1.915

[64] Rafael Sarkis-Onofre, Ferrán Catalá-López, Edoardo Aromataris, and Craig Lockwood. 2021. How to properly use the PRISMA Statement. *Systematic Reviews* 10, 1 (April 2021), 117. https://doi.org/10.1186/s13643-021-01671-z

[65] Leif Sigerson and Cecilia Cheng. 2018. Scales for measuring user engagement with social network sites: A systematic review of psychometric properties. *Computers in Human Behavior* 83, (2018), 87–105.

[66] Frans van der Sluis. 2025. Wanting information: Uncertainty and its reduction through search engagement. *Information Processing & Management* 62, 2 (March 2025), 103890. https://doi.org/10.1016/j.ipm.2024.103890

[67] Saira Hanif Soroya, Ali Farooq, Khalid Mahmood, Jouni Isoaho, and Shan-e Zara. 2021. From information seeking to information avoidance: Understanding the health information behavior during a global health crisis. *Information Processing & Management* 58, 2 (2021), 102440.

[68] Evropi Stefanidi, Marit Bentvelzen, PawełW. Woźniak, Thomas Kosch, Mikołaj P. Woźniak, Thomas Mildner, Stefan Schneegass, Heiko Müller, and Jasmin Niess. 2023. Literature Reviews in HCI: A Review of Reviews. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (*CHI '23*), April

19, 2023. Association for Computing Machinery, New York, NY, USA, 1–24. https://doi.org/10.1145/3544548.3581332

[69] Ayoung Suh, Christy MK Cheung, and Y. Q. Lin. 2019. Revisiting User Engagement: Concepts, Themes, and Opportunities. In *PACIS 2019 Proceedings*, 2019. AIS, 150. Retrieved from https://aisel.aisnet.org/pacis2019/150

[70] Mingfei Sun and Shaojing Sun. 2024. Users' engagement in WeChat-based support communities: A multilevel perspective. *Information Processing & Management* 61, 2 (March 2024), 103604. https://doi.org/10.1016/j.ipm.2023.103604

[71] Zhida Sun, Manuele Reani, Quan Li, and Xiaojuan Ma. 2020. Fostering engagement in technology-mediated stress management: A comparative study of biofeedback designs. *International Journal of Human-Computer Studies* 140, (August 2020), 102430. https://doi.org/10.1016/j.ijhcs.2020.102430

[72] Alistair Sutcliffe. 2000. On the effective use and reuse of HCI knowledge. *ACM Trans. Comput.-Hum. Interact.* 7, 2 (June 2000), 197–221. https://doi.org/10.1145/353485.353488

[73] Jeyan Thiyagalingam, Mallikarjun Shankar, Geoffrey Fox, and Tony Hey. 2022. Scientific machine learning benchmarks. *Nature Reviews Physics* 4, 6 (2022), 413–420.

[74] Yuan Tian, Ke Zhou, and Dan Pelleg. 2021. What and How Long: Prediction of Mobile App Engagement. *ACM Transactions on Information Systems* 40, 1 (September 2021). https://doi.org/10.1145/3464301

[75] Noam Tractinsky. 2018. The Usability Construct: A Dead End? *Human–Computer Interaction* 33, 2 (March 2018), 131–177. https://doi.org/10.1080/07370024.2017.1298038

[76] Noam Tractinsky. 2020. The Usability Construct: A Concern for Both Theory and Practice. *Human–Computer Interaction* 35, 4 (July 2020), 338–353. https://doi.org/10.1080/07370024.2019.1570466

[77] Raphael Velt, Steve Benford, and Stuart Reeves. 2020. Translations and Boundaries in the Gap Between HCI Theory and Design Practice. *ACM Trans. Comput.-Hum. Interact.* 27, 4 (August 2020), 1–28. https://doi.org/10.1145/3386247

[78] John G. Wacker. 2004. A theory of formal conceptual definitions: developing theory-building measurement instruments. *Journal of Operations Management* 22, 6 (2004), 629–650.

[79] Ben Wang and Jiqun Liu. 2023. Investigating the role of in-situ user expectations in Web search. *Information Processing & Management* 60, 3 (May 2023), 103300. https://doi.org/10.1016/j.ipm.2023.103300

[80] Kelly G. Wilson. 2001. Some notes on theoretical constructs: Types and validation from a contextual behavioral perspective. *International Journal of Psychology and Psychological Therapy* 1, 2 (2001), 205–215.

[81] Claes Wohlin, Per Runeson, Martin Höst, Magnus C. Ohlsson, Björn Regnell, and Anders Wesslén. 2012. Systematic Literature Reviews. In *Experimentation in Software Engineering*, Claes Wohlin, Per Runeson, Martin Höst, Magnus C. Ohlsson, Björn Regnell and Anders Wesslén (eds.). Springer, Berlin, Heidelberg, 45–54. https://doi.org/10.1007/978-3-642-29044-2_4

[82] John Wu, Yonatan Belinkov, Hassan Sajjad, Nadir Durrani, Fahim Dalvi, and James Glass. 2020. Similarity Analysis of Contextual Word Representation Models. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, July 2020. Association for Computational Linguistics, Online, 4638–4655. https://doi.org/10.18653/v1/2020.acl-main.422

[83] Wayne Wu. 2024. We know what attention is! *Trends in Cognitive Sciences* 28, 4 (April 2024), 304–318. https://doi.org/10.1016/j.tics.2023.11.007

[84] Bao-Jun Xia. 2025. Navigating user engagement and cultural transitions in entertainment technology and social media based on activity management. *Entertainment Computing* 52, (2025), 100791.

[85] Ruowei Xiao, Zhanwei Wu, and Juho Hamari. 2022. Internet-of-Gamification: A Review of Literature on IoT-enabled Gamification for User Engagement. *International Journal of Human–Computer Interaction* 38, 12 (July 2022), 1113–1137. https://doi.org/10.1080/10447318.2021.1990517

[86] Joanne Yu, Astrid Dickinger, Kevin Kam Fung So, and Roman Egger. 2024. Artificial intelligence-generated virtual influencer: Examining the effects of emotional display on user engagement. *Journal of Retailing and Consumer Services* 76, (2024), 103560.