DOI: 10.54254/2755-2721/2026.GU32164

Applications of Multimodal Technology in User Interfaces: A Systematic Review

Zhaohong Zhang

Department of Institute for Digital Technologies, Loughborough University London, London, UK
1150729480@qq.com

Abstract. Development of user interface (UI) is undergoing traditional grafic user interface to multi-modal and AI-driven paradigms , which significantly enhanced interaction richness and system adaptability. Traditional interface rely on interaction logic that predefined and offering limited flexibility in addressing diverse user behaviors, cognitive patterns, and contextual conditions. Advances in multimodal technologies and artificial intelligence recently provide new opportunities to overcome these limitations through supporting context-aware, adaptive, and personalized interaction. This work aims to exploring how multimodal interaction and AI-based methods affect the design, evaluation, and enhancement of UI/UX systems. Through using a systematic review approach, recent studies identifies key trends in multimodal presentation, AI-driven evaluation, and adaptive interface mechanisms. The findings illustrate a shift from a static, visually dominant interfaces toward intelligent systems in which AI functions as an evaluator, optimizer, and co-creator within iterative human–AI design workflows. This review is anaylzing recent research which published from 2018-2025 with keyword UI/UX and multi-modal. Besides,this paper provides a conceptual framework in order to guide the development of explainable, adaptive, and human-centered multimodal UI/UX systems.This study contributes more possibility for applicating and exploring in the interdisplinary of HCI and design.

Keywords: Multi-modal Technology, User Interface, User Experience, Artificial Intelligence, Large Language Model

1. Introduction

The progression observable within the domain of user interface design has manifested itself as a shift from conventional graphical paradigms toward configurations characterized by multi-modal integrativity. A comparably expanded scope and complication in modalities for human-computer interaction is thereby established, relative to antecedent methodological norms. Reliance, heretofore, on staticity and prescriptive logics in traditional interfaces circumscribed responsiveness primarily to narrow operational scenarios—seldom permitting efficacious accommodation of those cognitive idiosyncrasies, physio-behavioral singularities, or individuated contextual imperatives attributed to disparate users. It can be seen from recent advancements that trajectories instigated by emergent multi-modal technological innovation provide not only alternative perspectives but also potential resolutions vis-à-vis such adaptive insufficiencies inherent in prior systems.

This module aims to manage different informations like user interactions (e.g.,visualizations), text inputs (e.g., user commands), and individual data (e.g., browsing history), enabling the system to develop a deep adaptive understanding of user intentions and contexts. Through integrating user behavior preferences with real-time interaction data, the system can dynamically tailor the interface and interaction flow. Its core function is to create and adapt personalized"motions",which refers to interface behaviors and presentations,that can match users' requirements, preferences, and contexts accurately. This method allows the system to show the most appropriate information, options and controls, in order to promote usabiility, user engagement and operation efficiency. This adaptive user interface leverages multi-model inputs and interface adaptive technology to improve user experience in a variety of situations.

However, there is a gap between the potential of multimodal systems and their practical application in user interface design. At the mean time, it is unclear how to translate the contextual understanding and adaptive capabilities of LMMs into concrete, actionable design rules and ensuring a straightforward interaction and positive user experience, especially within complex socio-technical environments like Internet of Things (IoT).

Thus, this paper aims to answer question below:

RQ1: What influence the integration of multimodal technology bring to the design and evolution of user interface and user experience?

RQ2: What effects the adaptive interface driven by multimodal technology brings to user experience (UX) metrics, such as intuitiveness and cognitive load?

To investigate these questions and to propose the solution about multimodal adaptive mechanisms lack unified models and executable design specifications, this study will started the literature review based on the above research question. Meanwhile, we will use PRISMA guideline, search keyword including multimodal interaction, adaptive interface, user behavior modeling to find out 26 academic paper. Our research findings show that mutimodal significantly promote identify ability of users, but current research still meet some limitation such as insufficient real-time adaptation, weak cross-modal coordination mechanisms, and a lack of directly implementable design guidelines.

2. Conceptual background

2.1. The development of multi-model technology

Gilbert [1] introduced the concept of multi-modal technology, and explained it is an interaction pattern that operates through multiple modes beyond language alone, which includes logical, emotional, visceral, and intuitive forms. In this view, meaning and persuasion originate from language, emotion, and contextual signals, rather than just using linear linguistic reasoning. The earliest definition is seen as an inherent multimodality characteristic of human interaction, and highlighting that communication and understanding is formed through the integration of a variety of expressive modalities rather than a symbolic channel.

With development of technology, the defination of muti-modal has been changed. As Xv [2] said, the contemporary understanding do not seen modalities as separate, parallel input channels. Instead, multimodality define as a integrated processing and coordinated use of diverse heterogeneous data types, which including text, vision, audio, gesture, and contextual signals to form a unified and adaptive interpretation of user intent and environment. This core defination of new perspective including cross-modal representation learning, semantic alignment, and dynamic fusion. Based on these technique factors, recent review of literature already have a systematic organization, which forming a classical system of multimodal integration strategy, and also summarized latest context-awareness interaction [3]. These together achieve cross-modal to reason and tailor their responses accordingly. In human–computer interaction and intelligent systems area, multimodality has been seen as a foundation to form context-aware, personalized, and interactive interfaces. Current studies shows that, its meaning not only came from a single single channel in isolation, but from the relationships and interactions among modalities.

2.2. The effects on user interface/user experience

According to Ben Shneiderman [4], the defination of user interface (UI) originate from the principle of direct manipulation, which make the interface more transparent in the function and allows users more focus on their tasks. He emphasized that effective UI design should render system objects visible, and allows users to perform rapid, incremental, and reversible actions directly. Norman [5] was expand this view based on this defination later, he seen user experience (UX) as a emotion reflection and holistic perception that individual through product or system interaction, not only contain usability but also feelings, expectation and contextual meaning.

Nowadays, UX and UI have became a diversified interactive forms characterized by multimodal input and embodied interaction from statistic screen graphical interfaces. In early stages, it interacts through touch- and gesture-based, which is direct manipulation and multi-touch gestures that enabled intuitive control of digital content on mobile and tabletop devices [6]. After that, voice interaction was created as a complementary modality, which enables users to interact with the system through their voice, especially in hands-free and attention-constrained contexts such as smart assistants and in-vehicle interfaces. Recently, interface paradigms have been developed to augment virtual reality, and interactive situation gradually expanded to augmented and virtual reality,which has led to a breakthrough in interaction with spatial, embodied, and immersive environments, integrating gestures, speech, and spatial perception as a main function of the user interface [7]. This development illustrate that from UI and UX from pure visual effect to physical –digital experiences, in which context awareness and multimodal coordination play a central role in shaping adaptive and immersive interaction.

UI/UX development was at the same time as the development of multi-modal. Before that,interfaces rely on single-channel inputs, while rely on gesture, voice, sight, and physicial signial this period. This development represents an alteration from static interface presentation to more dynamic and adaptive interaction paradigms. With multimodal capabilities mature, UI/UX design was expanded to accommodate richer forms of interaction and greater contextual awareness, laying a clear developmental trajectory toward more responsive and personalized interface systems.

3. Methodology

This study will use a systematic review methodology to analyze empirical studies that are relevant to UI, UX, and multimodal design. Besides that,this study will focus on research that was published during the period from 2018 to 2025, in order to capture recent developments in User Interface (UI), User Experience (UX), and multi-modal interaction. Our research team form by three members, which advanced in computer science, design studies, and the psychology area, respectively which match PRISMA research standard to avoid a subjective error introduced by discussing a single topic. We conducted title/abstract screening on 2,333 articles according to the PRISMA standards, ultimately selecting overlapping articles identified by each team member to form a final dataset of 425 articles.

All of the searched articles originated from the ACM digital Library, IEEE, Scopus, and Web of Science database. These authority publication, like the International Journal of Human–Computer Interaction,The Design Journal, and so on, then,we supplemented the literature materials using keywords like intelligent interface systems.

During the process, we used the PRISMA guideline [8], and the initial search result was approximately 2753 records. After we expanded some repeated contents and non-related works that are not relevant to this area, 2333 articles was remained for title/abstract screening. Then,425 meet the relevance standard of concerning multimodaling behanvior data processing, motion system, or personalization. During the period, some paper will be excluded because of a focus on a single market or do not focus on the multi-modal or user interface area. Full-text assessment further reduced the pool to 253 studies. Then we will exclude some papers for the following reasons: non-general language, lack of design/technical innovation, and little user evaluation. Finally, there are only 26 document will be chosen (As shown in Figure.1).

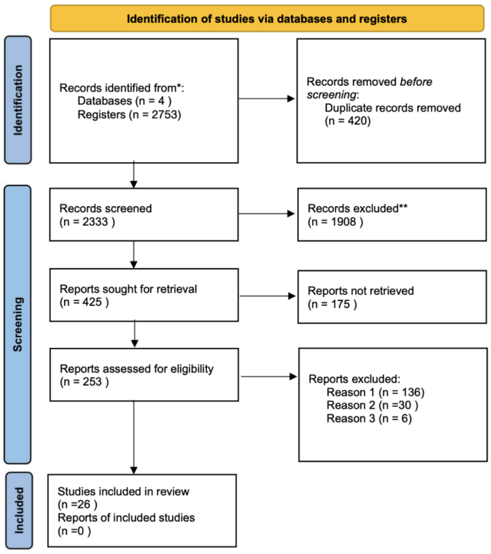


Figure 1. The flow diagram for systematic review

4. Results

During using the PRISMA method, we experienced the process of selecting studies across four key areas, which including interface presentation, the modes of interface presentation,mediatory functions assumed by AI, substratal data configuration, and employed technological implementations for multi-modal application. Meanwhile, there are within four different steps in our research, which built from the superficial exposition of feature emergences to more profound investigational penetration into model technicalities. According to this paradigm, following Results part will gradually illustrate insights, first will define and explain the basic phenomenological attributes and perceptible regularities, then we will layer by layer to explain and analysis of AI frameworks of the complex processes governed by the artificial intelligence framework and the methods employed in the related multimodal synthesis model. Through this allocate to form a hierarchical explication emerges, enabling understand systematic multi-modal UI&UX studies area current trend and innovation progress observed.

Table 1. The literature database of PRISMA inclusion

Study (Author)	Classification	Method	Task	Evaluation Metric
Faudzi et al. [9]	UI/UX Evaluation	Framework analysis & case-based evaluation	Mobile learning UI assessment	Usability criteria compliance
Iman et al. [10]	AI-Assisted UI Design	Prototype system & comparative user study	AI-supported minimalist UI design	User satisfaction, task efficiency
Namoun et al. [11]	UI/UX Evaluation	ML-based modeling & empirical validation	Usability prediction	Prediction accuracy, usability score
Park et al. [12]	AI-Assisted UI Design	Multimodal LLM system & user study	Inspirational UI search	Relevance rating, user preference
You et al. [13]	Multimodal UI Understanding	Vision-language modeling & benchmark testing	Mobile UI grounding	Grounding accuracy
Wu et al. [14]	Multimodal UI Understanding	Vision-language model & dataset evaluation	Intra-/inter-UI understanding	Retrieval and classification accuracy
Hussain et al. [15]	UI/UX Evaluation	Context-aware adaptive system & user experiment	Personalized UI adaptation	Task completion time, satisfaction
Oulasvirta et al. [16]	UI Optimization & Automation	Combinatorial optimization & simulation	GUI layout optimization	Objective cost reduction
Wu et al. [17]	UI Optimization & Automation	Multimodal LLM-based automation	UI-to-code generation	Code correctness, layout fidelity
Zhang et al. [18]	Safety, Accessibility & Ethics	Interaction impact modeling & empirical study	Mobile UI action risk analysis	Error rate, safety violations
Hilbert & Redmiles [19]	UI/UX Evaluation	Interaction log analysis & statistical modeling	Usability inference	Error frequency, usage patterns
Alomari et al. [20]	UI/UX Evaluation	Evaluation framework & validation study	Cyberlearning UX assessment	Usability and learning effectiveness
Chen et al. [21]	AI-Assisted UI Design	Autoencoder-based retrieval model	Wireframe-driven UI search	Retrieval similarity accuracy
Liu et al. [22]	Multimodal UI Understanding	Visual understanding model & fault localization	UI display issue detection	Precision, recall
Franco et al. [23]	UI/UX Evaluation	Multimodal sentiment analysis & visualization	UX state monitoring	Sentiment classification accuracy
Duan et al. [24]	UI/UX Evaluation	Dataset construction & automated critique	UI design quality assessment	Agreement with human judgment
Zhang et al. [25]	Interaction Techniques	Controlled user experiment	Multimodal interaction engagement	Engagement level, task performance
Minowa & Zhang [26]	Interaction Techniques	Hardware prototype & user testing	Pseudo-haptic interaction	Perceived realism
Pavlov et al. [27]	Safety, Accessibility & Ethics	Assistive system design & user study	Accessible mobile UI interaction	Task success rate
Wen [28]	UI/UX Evaluation	Visual design experiment	Mobile UI color design	Visual comfort, recognition accuracy

4.1. Presentation forms of multimodal technology in UI

In the current mobile UI system, traditional usability heuristics were still influenced, especially in consistency, visibility of system status, feedback, and error prevention. In order to solve these problems, these principles are always illustrated through simplified layouts, reduced interaction steps, and building clear visual hierarchies, which to adapt to mobile contexts for reason which avoid cognitive and motor load on small touchscreens [9]. Recently, there are some quantization methods to enhance UI layout complexity, balancing between information density and cognitive load [29]. Meanwhile, some multimedia design principles like signaling, redundancy reduction, and spatial contiguity are more and more applied to enhance users' engagement and form information flows. These strategies aim to enhance operation efficiency and cognitive load controlability, although meet some difficulty like limited screen and distraction audience attention, this strategy still can maintain usability of interface [10,28].

Nowadays, the mobile application became more and more complex, thus how to enhance interaction through sense judgment were became a important agenda of user interface (UI). Under this background, multimodal has gradually became a core situation of spread interaction requirement to each sensory channel. Current mobile interface more and more applied in speech input, auditory feedback, tactile cue and manual manipulation.These channel are supplement of of vision information to reduce stress of visual channel, thus stably and adaptive of interaction can be increase [26,27]. For example, voice recognition systems enable users finish navigation and order operation while auditory signal and vibration mode can provide confirm,warnings, or error feedback without visual attention. These multimodal presentation methods more inclusive when facing visual, motor and language impairments users,and match design idea [30].

Another trend that multi-modal present is that increasing context awareness and adaptive ability. Current user interface more focus on instant response quality of surroundings like illumination intensity, noise level, device status, user mobile status and network quality.When visual readability decrease, the system can intensify auditory feedback. When user is in motion, interaction flow also can adjust according to the situation [15]. Modal based adaptive framework switch modal focus according to integrate contextual information and judgement of user status, thus interaction process match actual situation.This adaptive represent a systemic transformationof interface from static layout to dynamic layout.

In recent year, the developement of UI has present multiple possible for multi-modal users, especially intelligent understanding and coprocessing of heterogeneous data sources.Visual language models (VLMs) and multimodal large language models (MLLMs) enable to analysis screenshot, layout, text describe and browser history at the same time to better understand interface semantics and user intention [13,14]. These capabilities support the automated analysis, evaluation, and generation of user interfaces, enabling the large-scale evaluation and optimization of multimodal presentation strategies [22,24]. Multimodal user interface presentation increasingly exhibits three significant characteristics: cross-modal cognitive load distribution, complementary and reinforced interaction channels, and dynamic environmental adaptability. These characteristics collectively lay the foundation for the integration of AI-driven evaluation, optimization, and generation methods, paving the way for smarter, more adaptive, and user-centric user interface systems [11,31].

4.2. Role of AI in UI/UX evaluation and optimization

In current UI/UX area, the usage of UI and large language modals (LLMs) is very common, which mainly role as automated assessment tool, design and optimize framework and interface simplification tools. In term of evaluation, traditional user interface relies on specialists' judgement and experience or small range experiment, which are gradually supplement or replaced by AI driven analysis method. AI can analysis massive interaction log, interface screenshot and layout to automate usability assessments and performance predictions. This process is more continuous but relies on dataset [11,14]. Visual language modal(VLM) has further enhance this ability, which not only can understand element of interface, navigation path, and cross-screen logic, and also support auto contrast and analysis from different version [13,14].

As for methodological level, framework based on AI and LLM is gradually integrating orginally scattered vision, text and behavior data which allocated them into a unified understanding and evaluation flow [13,14]. This process mode enable UI/UX analysis not relies on single dimension any more,but consider interface present,user behavior and contextual meaning.Furthermore, intervention of AI also enhance effiency of interface design and optimize. On one hand, adaptive and generative method enable designers explore in wider design space, provide conciseness, usage and layout placement scheme [10,15]. Combination optimization is a effective way to form and adapt GUI layout, which can balance technique requirement and user experience [16]. On another hand, automatic tool can alter interface design to executable code, reduce development code, shortage iteration cycle. In minimalism style, AI also have a good proformance which can effcient simplify interaction process and remove excessive visual elements [10].From a macro perspective, application of AI has promoting alternative of UI/UX search paradigm, current studies has exploring UI/UX methodology and how to apply AI tools,modals, and data into design and evacuate flow [11,32]. Application scenarios of multi-modal is maintain expanding. Moreover, it also cover complex task like UI code creation and cross-interface reasoning, shows a unique proformance when handle heterogeneous data [13,14].

4.3. Common characteristics of datasets in UI/UX research

In the recent year, AI and multi-modal develop rapidly. In this case, the crucial role of dataset in UI/UX is increasing prominent. Current mainstream dataset is widely use multi-modal structure, integrating various data type such as interface screenshot, layout information, visual feature and text highlight. This design enable learn interface appearance, interaction behavior and user intent of modal correlative cross between modals, thereby achieving a whole understanding and contextual awareness [11,13,14]. The value of dataset not reflected on simple statistical indicators, but measure by actual task proformance like the accuracy of usability prediction, the precision of interface understanding, the effectiveness of defect localization, and the quality of automatically generated evaluation text [22,24].

In guiding user interface design, these dataset enable to quantify present interaction result like task completion rate, navigation efficiency, error patterns, and visual clarity, which making the abstract usability concept become more measurable and actionable.Through structured expression towards user interface and interaction flow, AI modal can build a connection between configuration and user experience, thus predictive evaluation and design suggestion can be achieved [11,14]. Some key evaluation datasets can further link user interface function and qualitative evaluation criteria, which enable usability principles to embedded in automated evaluation processes [24].

Many datasets are collected from real situations, which reflecting the interactive behaviors of different real users in various application domains. These real-world-based datasets enable models to transcend the limitations of laboratory environments and support continuous, dynamic user interface optimization in real-world settings [22,25]. Through these case studies, we have identified the importance of datasets for integrating multimodal elements into user interface/user experience system design. In the long term, these characteristics will enable scalable evaluation, bridge the gap between design and behavior, and lay the foundation for large-scale user interface modeling. Such datasets are crucial for advancing the development of adaptive, intelligent, and human-centered user interface systems.

4.4. Multimodal approaches application in UI/UX

There are three key directions emerge from ongoing research into multimodal applications and user interface/user experience (UI/UX). First, interfaces enhance user interactivity and facilitate the allocation of cognitive resources by integrating visual, auditory, and behavioral metrics, thereby improving accessibility under the constraints of various contextual parameters. Second, artificial intelligence assumes an evaluative role, used for interface evaluation, alternative design generation, and optimization processes based on adaptive and data-driven criteria. Third, the context-awareness of interfaces is increasingly prominent; the importance of modality is thus dynamically adjusted based on environmental or personalized user variables—the resulting integration continuously ensures the consistency and personalization of the experience. The following sections will analyze typical cases to illustrate the transformation achieved in user interface design practice through the adoption of multimodal approaches.

4.4.1. Multimodal methods for supporting UI iteration

When we studying the application of multimodal technologies in user interface/user experience design, we can observe three key trends. A significant alternative is that combination visual, auditory, and behavioral that can enhance users' engagement, and redistribute cognitive load and diffuse infrastructure across diverse contexts, leading to improve its accessibility. Artifical intelligence is using on enhance evaluate of interface and creation. Its analysis ability can also use to evaluate operate and propose different design schemes, which improving adaptive mode of highly response data, and improve the process and make it more iterative in nature. These interface shows strongly scene perception ability, Modal adaptation closely monitors context or individual user behavior to deliver a consistent yet distinctly personalized user experience. Subsequent chapters will focus on analyzing typical cases to demonstrate the transformative impact of adopting a multimodal approach in contemporary user interface architecture paradigms.

In the past, the development of visual and large scale language models (LLMs) is further improve the user interaction evolution, which make the model can understand interface componements and navigation process. These modes support automated user interface evalution, design contrast and cross-screen reasoning.Compared to traditional modal models, LLM achieves a shift from single-modal understanding to cross-modal reasoning, enabling designers to efficiently evaluate multiple design options [13,14]. Furthermore, the automated approach to multimodal user interfaces (UIs) can directly translate layout information and visual structure into executable UI code, significantly shortening the feedback cycle between design and implementation [17]. Through capture UI layout like element reletive and layout tree, new framework enable to restore originate layout when new code is forming [28]. This capability enables rapid prototyping and iterative experimentation without extensive manual redevelopment.

Multimodal method combine real users behavior with design process, support iteration impovement. Interaction logs, behavioral sequences, and performance trajectories will be recorded to provide empirical evidence of how users interact with UI elements, while quantitative assessments and physiological measurements capture cognitive load and emotional responses during interaction, enabling AI systems to correlate design features with usability outcomes [22,25]. This data-driven iteration facilitates informed decision-making and reduces subjective bias in design evaluation. In summary, the multimodal approach transforms user interface iteration into a more systematic, efficient, and scalable process, laying the foundation for an intelligent, adaptive interface design workflow.

4.4.2. Theoretical innovations supporting multimodal applications in UI

These multimodal methods are becoming more and more popular in UI/UX design, this was from a series of theoretical innovations and it expanded traditional interaction and usability models. While traditional UI theories relies on visual perception, task efficiency, and heuristic-based evaluation, and seen user and system as a process of single dominant channel to dispose messages. However, recent year studies shows that user interface is a process of multimodal cognitive. During this process, users to perceive, understand and operate interface through multiple sense like visual, language, auditory and tactile [31,32].

There is a important findings is integrating multimodal perception and cognitive theories in user interface framework. This theories emphasis load batching of different perception channel,and believe that effective user interface should balancing visual, auditory and tactile rather than put all of cognitive stress in single channel. This insight provides a effective theoretical support for multimodal interface strategy, especially in complex environment and mobile devices, which good for reduce cognitive overload and enhance user performance [9,25].

In AI centered interface theories area also achieved a breakthrough like large scale user interface framework and data driven user experience framework. These theories altered UI/UX design from rule-based heuristic approach to study base system, enabling infer usability patterns and interactions modals from massive multimodal dataset [33]. Through formed interface elements to structured and learnable formal expression, these modals established a foundation for automatic evaluation, adaptive interaction and generative design. Meanwhile, interaction machine learning has further emphasis two-way interactive relationship between user and AI, and established transparency, controllability, human- computer interaction as a foundation of multimodal interface application [34].

In summary, these theoretical innovations provides us theoretical support for understanding how to coordinate multimodal signal in interface system, and also effective linked human-centered design idea and AI driven technological path, established a foundation for sustainable develop to create intelligent, adaptive interface.

4.4.3. Comparative analysis of multimodal interaction methods and their relative advantages

In multimodal user interface/user experience (UI/UX) research, different modal combinations exhibit significant differences in information delivery efficiency, cognitive load control, and user experience quality. Current studies mainly about three type of interaction form to compare and analysis: single mode like pure text and visual,or bi-modal like visual and language , visual and auditory, or multimodal mixture like visual, auditory and tactile.Through these studies, it shows different combination method their interaction advantage and applicable scenarios, its effects depend on specific task environment and user situation [11,32].

In traditional user interface, single visual mode occupied dominant for long time and widely used because of its clear structure, high efficiency to show complex information. In user interface and user experience area, information density defined information carried in limit space through visual element like icon, text and layout. This presentation form good for users to understand content in high efficiency but it easy to distract and reduce user decision accuracy and interface usability [9]. In order to solve this problem, designers always use image-text method to enhance information clarity and language meaning expression, which is more and more used in user interface and automated design in current large language modal support [14]. However, this compound mode cannot leave without users' dependence of reading ability and visual processing ability. This limitation shows that it it necessary to explore cross-modal scheme especially auditory vision fusion mechanism to effective allocate cognitive load and maintain user participation .

Based on existing research, audiovisual multimodal fusion exhibits the highest relative advantages in usability, task completion efficiency, and user engagement. This approach effectively distributes cognitive load by presenting structured information through the visual channel while providing immediate feedback, cues, or status updates through the auditory channel. This cross-modal coordination mechanism significantly accelerates user perception of interface changes, particularly suitable for mobile devices, complex workflows, and real-time interactive scenarios [13,25]. In user interface/user experience (UI/UX) design practice, this design manifests as an interface that combines visual cues with auditory elements such as alerts, confirmations, or guidance prompts, including notifications, progress indicators, and interactive layouts. This integration ensures that users can more intuitively monitor and respond to interface changes, thereby improving overall interaction quality and user satisfaction.

Furthermore, audiovisual interaction offers significant advantages in accessibility design and inclusive experiences. Auditory feedback can compensate for users with limited visual attention or visual impairments, reducing reliance on sustained visual attention and enhancing the robustness and adaptability of the interface. Compared to multimodal solutions incorporating haptic feedback, audiovisual systems are more feasible in terms of hardware dependence, implementation costs, and cross-platform deployment, making them more valuable in practical UI/UX applications [35]. In UI/UX practice, this combination typically manifests as an interface that integrates visual elements such as icons, progress bars, or text prompts with auditory cues such as alerts, confirmations, or voice commands. This fusion not only supports inclusive interaction but also ensures that users efficiently perceive and respond to interface changes, thereby improving overall usability and user engagement. In conclusion, compared to other modal combinations, audiovisual interaction demonstrates a more balanced and significant advantage in cognitive efficiency, user engagement, system scalability, and practical feasibility. This makes it the most viable interaction paradigm for AI-driven multimodal user interface/user experience optimization.

Overall,visual-auditory interaction have more significant advantage compared to other modality combinations in cognitive efficiency, user engagement, system scalability, and practical application feasibility. This establishes it as the most actionable interaction paradigm for AI-driven multimodal UI/UX optimization.

5. Discussion

When we expanding our findings, we found that users' perception and interaction with user interfaces often face the risk of cognitive load, and define it as a mental effort that to dispose information and finish task, this in the range of people to work and memory. This research highlight the importance of the multimodal interaction to improve user interface (UI) and user experience (UX) design. Althrough traditional usability is still important, but it is supplement by more and more multimodal method, and these method improve accessibility and help distribute cognitive load across multiple sensory channels. Through combine visual, auditory, tactile and behavior model, enabling user interaction more nature,and better adaptive different ability and environmental background. Artificial intelligence has altered this trend, which can automated design improve and availability assessment. This method that driven by AI combined with human-in-the-loop (HITL), support sustaining and evidence-based UI/UX improvement. This studies propose an AI-mediated framework for multimodal UI/UX that integrates sensing, AI learning, and adaptive interface behaviors. However, there are some risks like limit multi-modal dataset and the opacity of AI decision-making remain, future work must solve these problems to create more transparent, ethical, and inclusive systems.These development lead to more adaptive, data driven UI/UX paradigm, which evolves through sustained human–AI collaboration.

5.1. Summary and elevation of findings

The findings of this study reveal a multi-layered understanding of how multimodal and AI-driven technologies are transforming UI/UX research and design, and there are four key insights: Firstly, multi-modal presenation form can disperse distribute cognitive load to different sensory channels to improve accessibility and adaptive interaction. Second, AI and large language scale modals role as evaluative and generative agents which automate usability assessment and design optimization. Third, multimodal datasets have become the infrastructure for scalable, data-driven evaluation and learning. Fourth, theoretical and practical of multi-modal interaction framework further present how AI integrates perception, cognition, and design iteration into unified workflows. To sum up, these findings shows a evolutionary trajectory from static, visually dominated interfaces toward intelligent, adaptive, and user-centered ecosystems.

Recent findings of AI and multimodal research has further reinforced the availability and importance of these findings.Some studies such as Ferret-UI [13] and MobileVLM [14] demonstrate multimodal logical hierarchy models (LLMs) can understand user interface semantics, interaction intents, and layout logic to support automated evaluation, generation, and reasoning. Meanwhile, Namoun et al. [11] and Iman et al. [10] prove that large-scale user interface models and minimalist AI frameworks through balance automation and interpretability to improve design efficiency and sustainability. Besides,Stige et al. [31] and Dudley and Kristensson [32] also emphasize that integrating AI into user experience evaluation, which not only keep human-centered design principles but also expanding the scale of analysis and predictive accuracy.These studies confirm that our studies have theoretical robustness and technical practicality although under the background of latest multimodal intelligence paradigm.

According to these insights, these studies results have transformative potential in multiple fields. In education area, adaptive multimodal interfaces can enhance accessibility and cognitive engagement; in healthcare area, context-aware user interface systems can enable personalized patient interactions. Furthermore, in the creative industries area, AI-driven generative frameworks can accelerate creative brainstorming and prototyping. This research is driving a new paradigm in user interface/user experience design gradually,which evolving from human-computer interaction to human-computer collaborative creation, thereby building responsive, inclusive, and intelligent digital ecosystems.

5.2. Intelligent interface convergence & evolution framework

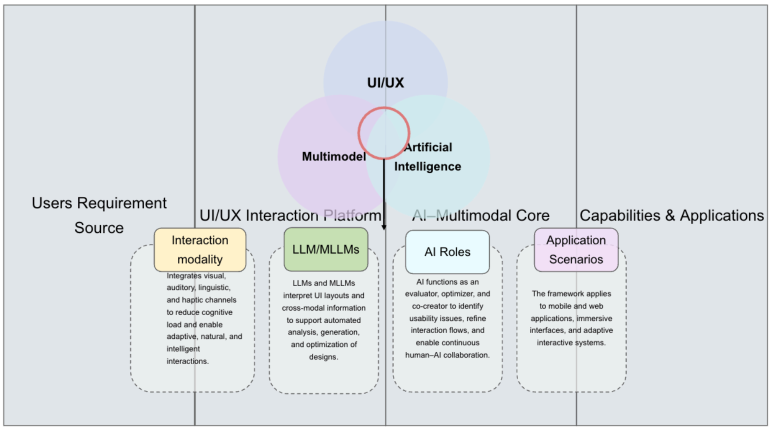
The framework form a integrated model by integrate multi-layered findings to explain multi-modal interaction, artificial intelligence and large language models (LLMs) how to reshape design, evaluate and optimization of user interface (UI). Previous studies of user interface emphaize that design of smart system should be clearity character allocate, adaptive automatic extend and continuous user control to improve human capabilities [36]. Expanding these priniciple to form AI and LLM-based systems, recent studies describe this type of modal a new interface layer which can natural language interaction, adaptive content generation, and dynamic user assistance, and also it bring some risks about transparency, controllability, and user trust [37]. Thus, this framework (as shown in Figure.2) work around four interdependent dimensions—interaction modality, large language models, AI roles, and application scenarios. Each of them represent a dynamic subsystem within the broader ecosystem of intelligent UI/UX design.

This framework below is based on multi-modal interaction, which surpassed traditional visual-tactile design, enabling the coordinated sensory engagement. Visual, auditory, linguistic,and tactile modalities are cooperated to distribute cognitive load, enhance feedback and improve possibility.As for current studies, these system integrate multiple input/output channels to enhance usability and straight forward cross-context [3]. Multimodal interface use complementary sensory channels to improve adaptive and robustness, which enabling this system to prioritize modals according to environmental conditions [38]. Context-aware adaptation allows interface respond to changing environmental factors and user condition to adjust modal emphasis and interaction strategies, maintain consistency and inclusivity. Recent studies shows this ability is very important for adaptive and smart user interface [30]. This coodination altered interface as a adaptive communication medium to learning and respond users rather than keep maintain. This undoubted align with the human-computer interaction widely trend, which is systems integrate multimodal inputs and adapt to context in real time.

Then the core of the second layers goes to integration of LLMs and multimodal LLMs (MLLMs), which role as semantic engines for UI understanding and creation. These modal able to explain and understand by screenshots, layouts, interaction flows, and textual metadata to get high-level semantic representations of design intent. Through cross-modal reasoning, LLMs can automatic critique, UI-to-code translation, and generative redesign. These role as a bridge to explain, link the symbolic reasoning of human designers with the perceptual reasoning of AI systems, accelerating iteration and ensuring semantic coherence across modalities.

The third dimension seen AI as the partner during the process of analytical and creative partner in UI/UX workflows. During this period, AI will role as a evaluator through multi-modal reasoning and analysis usability issues and predicting performance. As a optimizer, it improve layout and interaction flows through adaptive of modal-base, while role as a cooperator,create a new design alternatives that aligned with user behavior and contextual data.These complementary roles build a continuous feedback circulation between human professional knowledge and computational intelligence. Promoting data-driven design decision that both scalable and human-centered.

The final dimension put the framework in diverse real-world areas. In mobile and wearable systems, multimodal AI supports responsive efficiently and low-attention interactions; in education, adaptive interfaces enable personalize learning and enhance engagement; in healthcare, context-sensitive UIs facilitate accessible and empathetic communication; and in creative industries, generative models assist rapid prototyping and aesthetic exploration. Across these domains, the framework promotes the design of systems that are perceptually rich, contextually intelligent, and aligned with human values.



Integrates visual, auditory, linguistic, and haptic channels to reduce cognitive load and enable adaptive, natural, and intelligent interactions.

LLMs and MLLMs interpret UI layouts and cross-modal information to support automated analysis, generation, and optimization of designs.

AI functions as an evaluator, optimizer, and co-creator to identify usability issues, refine interaction flows, and enable continuous human–AI collaboration.

The framework applies to mobile and web applications, immersive interfaces, and adaptive interactive systems.

Figure 2. A conceptual framework of AI-driven multimodal UI/UX interaction platform

5.3. Limitation

Although the framework provides a comprehensive perspective on the integration of multimodal interaction and AI-driven intelligence in UI/UX design, it still faces several intrinsic limitations related to technological maturity, data dependence, and contextual adaptability. These constraints are partly unavoidable given the current stage of multimodal AI development and the evolving nature of human–AI interaction research.

Although this study emphasizes user-centered adaptability, the proposed framework does not fully capture the emotional, cultural, and situational nuances that fundamentally shape user experience.Current multimodal analytics primarily quantify observable user behavior, which limits their ability to interpret subjective perception, emotional response, and aesthetic judgment, potentially resulting in an incomplete understanding of user satisfaction. Furthermore, users with limited digital literacy, as well as those operating in non-visual, low-interaction, or resource-constrained environments, may not benefit equally from multimodal and AI-enhanced interfaces. In cross-cultural contexts, identical information, symbols, or visual representations can lead to markedly different semantic interpretations across regions, ethnic groups, and language communities [39]. These challenges highlight the continuing necessity of human oversight, participatory design, and ethical review, as well as the need to develop more inclusive and generalizable cross-cultural UI/UX design guidelines that minimize misinterpretation while respecting cultural diversity [40].

The limitations of this framework are also affected primarily by the current technological boundaries of multimodal learning, data availability, and contextual variability. Addressing these challenges requires advances in explainable AI, inclusive dataset design, and real-world evaluation methodologies that bridge computational intelligence with diverse human needs.

5.4. Future work

Future research should aim to refine and expand the proposed framework by integrating human-centered design principles, explainable AI, and cross-domain multimodal intelligence. One immediate direction lies in developing more diverse and ethically curated multimodal datasets that capture a broader spectrum of user behaviors, contexts, and accessibility needs. Such datasets would enhance the inclusivity and generalization capacity of AI-driven UI systems, mitigating current limitations related to bias and contextual adaptation.

Besides,the oretical enrichment remains essential. Future studies should formalize the cognitive and perceptual foundations of multimodal interaction, creating unified evaluation metrics that connect usability, emotion, and ethics. Through these advancements, the framework could evolve into a comprehensive human-AI symbiotic model—one that not only enhances UI/UX intelligence but also redefines the boundaries of digital creativity, accessibility, and adaptive interaction in the age of multimodal AI.

In addition, expanding the framework toward real-time adaptive systems represents an important step forward. Future research should investigate lightweight and edge-based implementations that bring multimodal reasoning to mobile, wearable, and ambient computing environments. These directions could enable more fluid, context-aware, and sustainable design workflows that extend beyond laboratory conditions into real-world scenarios.

Lastly, another promising avenue involves advancing explainable and interactive multimodal models that enable designers to interpret, question, and adjust AI-generated insights. By embedding transparency and interpretability mechanisms within large language and multimodal models, future systems could establish greater trust and accountability between human designers and computational agents. This aligns with emerging efforts in human-AI co-design—a paradigm that envisions designers not as passive recipients of algorithmic recommendations but as active collaborators in adaptive, iterative design loops.

6. Conclusion

This paper discussed the evolution of user interface (UI) and user experience (UX) design from traditional graphical interface to AI driven system, focus on adaptive interaction and smart support. Through base PRISMA systematic literature review, this paper find that user interface is development towards respond user behavior, environmental and cognitive state. Research result illustrate the value of multimodal interaction in reducing cognitive load, improving accessibility, and increasing robustness, particularly in mobile and complex settings. AI especially large language and LLMs has became a key factor that promote scalable evaluation, iterative design, and generative optimization, augmenting human designers, through data driven observe to improve human design ability. Base the result and discussion, this paper proposed a unified conceptual framework which integrated multi-modal design, artificial intelligence characters and application situation to support adaptive and human-centered design.Althrough there are still have some challenges on data diversity, interpretability, and ethical oversight, this paper build a foundation for create a adaptive, transparent, and aligned with human values UI/UX systems in the future.

References

[1] Gilbert, M. A. (1994). Multi-modal argumentation. Philosophy of the Social Sciences, 24(2), 159--177. https: //doi.org/10.1177/004839319402400202

[2] Xv, Guipeng, et al. (2025). "Unveiling the Impact of Multi-modal Content in Multi-modal Recommender Systems." Proceedings of the 33rd ACM International Conference on Multimedia. https: //doi.org/10.1145/3746027.3755300

[3] Dritsas, E., Trigka, M., Troussas, C., & Mylonas, P. (2025). Multimodal interaction, interfaces, and communication: A survey. Multimodal Technologies and Interaction, 9(1), 6. https: //doi.org/10.3390/mti9010006

[4] Shneiderman, Ben. (1983). "Human factors of interactive software." IBM Germany Scientific Symposium Series. Berlin, Heidelberg: Springer Berlin Heidelberg. https: //doi.org/10.1007/3-540-12273-7_16

[5] Norman, D. A. (1998). The invisible computer: Why good products can fail, the personal computer is so complex, and information appliances are the solution. MIT Press.

[6] Norman, Don. (2013). The design of everyday things: Revised and expanded edition. Basic books.

[7] Milgram, P., & Kishino, F. (1994). A taxonomy of mixed reality visual displays. IEICE Transactions on Information and Systems, 77(12): 1321--1329. https: //doi.org/10.1587/e77-d.12.1321

[8] Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ, 372, n71. https: //doi.org/10.1136/bmj.n71

[9] Ahmad Faudzi, M., Che Cob, Z., Omar, R., Sharudin, S. A., & Ghazali, M. (2023). Investigating the user interface design frameworks of current mobile learning applications: A systematic review. Education Sciences, 13(1), 94. https: //doi.org/10.3390/educsci13010094

[10] Iman, J. A., Felisha, A., Kimeison, M., Hermawan, E. S., Rumagit, R. Y., & Pranoto, H. (2025). Refining UI/UX with minimalist design and AI: Towards sustainable and efficient digital experiences. Procedia Computer Science, 269, 669--680. https: //doi.org/10.1016/j.procs.2025.01.089

[11] Namoun, A., Alrehaili, A., Nisa, Z. U., Almoamari, H., & Tufail, A. (2024). Predicting the usability of mobile applications using AI tools: The rise of large user interface models, opportunities, and challenges. Procedia Computer Science, 238, 671--682. https: //doi.org/10.1016/j.procs.2024.05.063

[12] Park, S., Song, Y., Lee, S., Kim, J., & Seo, J. (2025). Leveraging multimodal LLM for inspirational user interface search. Proceedings of the ACM on Human-Computer Interaction, 9(CSCW1), Article 359, 1--22. https: //doi.org/10.1145/3706598.3714213

[13] You, K., Chen, X., Liu, Y., Zhang, Y., Li, Z., Zhu, L., ... & Liu, Z. (2025). Ferret-UI: Grounded mobile UI understanding with multimodal LLMs. In A. Leonardis, E. Ricci, S. Roth, O. Russakovsky, T. Sattler, & G. Varol

(Eds.), *Computer Vision -- ECCV 2024* (Lecture Notes in Computer Science, Vol. 15122, pp. 245--262). Springer, Cham. https: //doi.org/10.1007/978-3-031-73039-9_14

[14] Wu, Q., Xu, W., Liu, W., Tan, T., Liujianfeng, L., Li, A., ... & Shang, S. (2024, November). Mobilevlm: A vision-language model for better intra-and inter-ui understanding. In Findings of the Association for Computational Linguistics: EMNLP 2024 (pp. 10231--10251). Association for Computational Linguistics.

[15] Hussain, J., Ul Hassan, A., Muhammad Bilal, H. S., Ali, R., Afzal, M., Hussain, S., ... & Lee, S. (2018). Model-based adaptive user interface based on context and user experience evaluation. Journal on Multimodal User Interfaces, 12(1), 1--16. https: //doi.org/10.1007/s12193-017-0252-2

[16] Oulasvirta, A., Dayama, N. R., Shiripour, M., John, M., & Karrenbauer, A. (2020). Combinatorial optimization of graphical user interface designs. Proceedings of the IEEE, 108(3), 434--464. https: //doi.org/10.1109/JPROC.2020.2969687

[17] Wu, F., Gao, C., Li, S., Wen, X. C., & Liao, Q. (2025). MLLM-Based UI2Code automation guided by UI layout information. Proceedings of the ACM on Software Engineering, 2(ISSTA), 1123--1145. https: //doi.org/10.1145/3728847

[18] Zhang, Z., Schoop, E., Nichols, J., Mahajan, A., & Swearngin, A. (2025, March). From interaction to impact: Towards safer AI agent through understanding and evaluating mobile UI operation impacts. In Proceedings of the 30th International Conference on Intelligent User Interfaces (pp. 727--744). Association for Computing Machinery. https: //doi.org/10.1145/3708359.3712075

[19] Hilbert, D. M., & Redmiles, D. F. (2000). Extracting usability information from user interface events. ACM Computing Surveys, 32(4), 384--421. https: //doi.org/10.1145/371578.371594

[20] Alomari, H. W., Ramasamy, V., Kiper, J. D., & Potvin, G. (2020). A user interface (UI) and user experience (UX) evaluation framework for cyberlearning environments in computer science and software engineering education. Heliyon, 6(5), Article e03945. https: //doi.org/10.1016/j.heliyon.2020.e03945

[21] Chen, J., Chen, C., Xing, Z., Xia, X., Zhu, L., Grundy, J., & Wang, J. (2020). Wireframe-based UI design search through image autoencoder. ACM Transactions on Software Engineering and Methodology, 29(3), Article 16, 1--31. https: //doi.org/10.1145/3387114

[22] Liu, Z., Chen, C., Wang, J., Huang, Y., Hu, J., & Wang, Q. (2022). Nighthawk: Fully automated localizing UI display issues via visual understanding. IEEE Transactions on Software Engineering, 49(1), 403--418. https: //doi.org/10.1109/TSE.2022.3145605

[23] Franco, R. Y. D. S., De Freitas, A. A., Lima, R. S. D. A. D., Mota, M. P., Dos Santos, C. G. R., & Meiguins, B. S. (2019, July). UXmood: A tool to investigate the user experience (UX) based on multimodal sentiment analysis and information visualization (InfoVis). In 2019 23rd International Conference Information Visualisation (IV) (pp. 175--180). IEEE. https: //doi.org/10.1109/IV.2019.00033

[24] Duan, P., Cheng, C. Y., Li, G., Hartmann, B., & Li, Y. (2024, October). UICrit: Enhancing automated design evaluation with a UI critique dataset. In Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology (pp. 1--17). Association for Computing Machinery. https: //doi.org/10.1145/3654777.3676395

[25] Zhang, L., Yu, J., Zhang, S., Li, L., Zhong, Y., Liang, G., ... & Lan, Z. (2024). Unveiling the impact of multi-modal interactions on user engagement: A comprehensive evaluation in AI-driven conversations. arXiv preprint. https: //arxiv.org/abs/2406.15000

[26] Minowa, H., & Zhang, C. (2020, September). Development of Touch Valve UI with pseudo-haptics feedback based on vibration of tablet PC. In *2020 9th International Congress on Advanced Applied Informatics (IIAI-AAI)* (pp. 456--462). IEEE. https: //doi.org/10.1109/IIAI-AAI50429.2020.00088

[27] Pavlov, N., Castro, M., Chukanska, Y., Pérez, C. M., Mileva, N., & Albert, M. J. (2018, October). Mobile graphical user interface with people with verbal communication disorders. In 2018 IEEE 5th International Congress on Information Science and Technology (CiSt) (pp. 391--395). IEEE. https: //doi.org/10.1109/CIST.2018.8656613

[28] Wen, G. (2021, August). Research on color design principles of UI interface of mobile applications based on vision. In 2021 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA) (pp. 539--542). IEEE. https: //doi.org/10.1109/AEECA52812.2021.00095

[29] Akça, E., Tanrıöver, Ö. Ö., & Yılmaz, A. E. (2025). A multi-criteria perceptual visual complexity metrics based optimization approach to user interface design layouts. IEEE Access, 13, Article 108420. https: //doi.org/10.1109/ACCESS.2025.108420

[30] Todorov, T., & Dochkova-Todorova, J. (2023, October). Accessible UX/UI design. In 2023 International Conference Automatics and Informatics (ICAI) (pp. 362--366). IEEE. https: //doi.org/10.1109/ICAI586213.2023.00068

[31] Stige, Å., Zamani, E. D., Mikalef, P., & Zhu, Y. (2024). Artificial intelligence (AI) for user experience (UX) design: A systematic literature review and future research agenda. Information Technology & People, 37(6), 2324--2352. https: //doi.org/10.1108/ITP-10-2022-0858

[32] Dudley, J. J., & Kristensson, P. O. (2018). A review of user interface design for interactive machine learning. ACM Transactions on Interactive Intelligent Systems, 8(2), Article 8, 1--37. https: //doi.org/10.1145/3185513

[33] Kristić, M., Zakarija, I., Škopljanac-Mačina, F., & Car, Ž. (2025). Machine learning for adaptive accessible user interfaces: Overview and applications. Applied Sciences, 15(23), Article 12538. https: //doi.org/10.3390/app152312538

[34] Holzinger, A. (2016). Interactive machine learning for health informatics: When do we need the human-in-the-loop? Brain Informatics, 3(2), 119--131. https: //doi.org/10.1007/s40708-016-0042-6

[35] Ramtohul, A., & Khedo, K. K. (2025). Adaptive multimodal user interface techniques for mobile augmented reality: Frameworks, modalities and user interaction. Array, 27, Article 100487. https: //doi.org/10.1016/j.array.2025.100487

[36] Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., ... & Horvitz, E. (2019). Guidelines for human-AI interaction. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (Paper 3, pp. 1--13). Association for Computing Machinery. https: //doi.org/10.1145/3290605.3300233

[37] Weisz, J. D., He, J., Muller, M., Hoefer, G., Miles, R., & Geyer, W. (2024, May). Design principles for generative AI applications. In Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (pp. 1--22). Association for Computing Machinery. https: //doi.org/10.1145/3613904.3642596

[38] Hu, Y. O., Tang, J., Gong, X., Zhou, Z., Zhang, S., Elvitigala, D. S., & Quigley, A. J. (2025, April). Vision-based multimodal interfaces: A survey and taxonomy for enhanced context-aware system design. In Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (pp. 1--31). Association for Computing Machinery. https: //doi.org/10.1145/3706598.3714161

[39] Marcus, A. (2009). Global/intercultural user interface design. In A. Sears & J. A. Jacko (Eds.), Human-Computer Interaction: Design issues, solutions, and applications (pp. 45-70). CRC Press.

[40] Smith, A., & Yetim, F. (2004). Global human--computer systems: Cultural determinants of usability. Interacting with Computers, 16(1), 1--5. https: //doi.org/10.1016/j.intcom.2003.11.001