Human-AI-UI Interactions Across Modalities

KEWEN PENG, University of Utah, United States
JEFFREY NICHOLS, Apple Inc., United States
CHRISTOF LUTTEROTH, University of Bath, United Kingdom
TIFFANY KNEAREM, MBZUAI, United Arab Emirates
FELIX KRETZER, Karlsruhe Institute of Technology (KIT), Germany
JEFFREY BIGHAM, Carnegie Mellon University & Apple Inc., United States
ALEXANDER MAEDCHE, Karlsruhe Institute of Technology (KIT), Germany
YUE JIANG, University of Utah, United States

Designing and developing user-friendly interfaces has long been a cornerstone of HCI research, yet today we are at a turning point where UIs are no longer designed solely for humans but also for intelligent agents that act on users' behalf, while UIs are also expanding beyond 2D screens into extended reality environments with inherently multimodal characteristics, together challenging us to rethink the role of the UI as a mediator of human–AI interaction. This workshop will explore how UI agents bridge human intent and system behavior by interpreting multimodal inputs and generating adaptive outputs across surfaces from screens to extended reality (XR), and we will examine not only their technical capabilities but also their broader impact, including how agents reshape daily workflows, how bidirectional alignment between human and AI activity can be achieved, and how generative models may transform UI creation. XR provides a compelling testbed for these questions and highlights challenges around accuracy, efficiency, transparency, accessibility, and user agency, setting the stage for the next generation of intelligent and adaptive UIs.

ACM Reference Format:
Kewen Peng, Jeffrey Nichols, Christof Lutteroth, Tiffany Knearem, Felix Kretzer, Jeffrey Bigham, Alexander Maedche, and Yue Jiang. 2026. Human-AI-UI Interactions Across Modalities. In *Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems (CHI EA '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3772363.3778737

1 OVERVIEW

The integration of artificial intelligence (AI) into user interfaces (UIs) is reshaping how people design, interact with, and experience digital systems. While traditional UI design has emphasized principles from cognitive psychology, such as mental models, affordances, and minimizing the gulfs of execution and evaluation [76, 81], the emergence of AI introduces new possibilities and challenges. Today, UIs are not only designed for human users but also for intelligent agents that act on users' behalf, while interfaces increasingly extend beyond 2D screens into immersive XR environments with inherently multimodal characteristics. These developments compel a rethinking of the UI as a mediator of human–AI interaction.

Recent advances in AI, particularly large-scale multimodal models, enable interfaces that can interpret diverse inputs, generate adaptive outputs, and learn from user behavior. Datasets such as Rico [7] and WebUI [102] have supported deep learning approaches for layout synthesis [26, 34, 35, 43, 46, 47, 62, 72, 85], code generation [71, 101], semantic modeling [1, 63, 107], and predicting user behavior [14, 36, 38, 39, 87, 90]. These capabilities open opportunities for more adaptive, generative, and collaborative interfaces while maintaining human oversight [13, 54, 55, 93, 100].

Alongside technical advances, AI is emerging as a partner in design and interaction workflows. Designers leverage AI for inspiration, generative explorations, and automating repetitive tasks [18, 28, 51, 71], while new research examines how AI can facilitate collaboration across designers, developers, and end users [66]. However, these possibilities raise critical questions about preserving human agency, trust, and control [51, 73].

Graphic UI (GUI) agents and adaptive XR interfaces exemplify these trends. In this proposal, we define GUI agents as autonomous, multimodal AI assistants that perceive interface state, interpret user intent, and act on the user's behalf within GUI from 2D interface to XR platforms. Powered by large language and vision–language models, these agents can perform complex tasks across web, mobile, and OS environments [9, 22, 103, 106, 108, 113]. When extended to XR settings, these interfaces increasingly leverage context-aware methods to optimize spatial layouts, content placement, and user engagement [4, 28, 56, 68, 88]. Integrating generative AI further supports mixed-initiative, personalized, and adaptive interactions, bridging user intent and system behavior across multimodal inputs and outputs [37, 64, 75, 96].

Despite these advances, core challenges remain unresolved. Designing AI-driven UIs requires balancing automation with human agency, achieving bidirectional alignment between human and AI activity, and enabling trustworthy, ethical, and inclusive interaction. This workshop therefore focuses on GUI agents as autonomous partners during use, multimodal interaction techniques for adaptive UIs, and computational interfaces that operate across modalities. Our goal is to build a cohesive research agenda for how intelligent interfaces can collaborate with users, rather than replace them, through situated, multimodal, and adaptive behaviors. To realize this goal and maximize long-term impact, we will produce a set of concrete post-workshop deliverables, including an open-access computational GUI agent research agenda, community channels for shared resources, and opportunities for collaborative publications on future human–AI–UI integration (see Section 8).

We have successfully organized four CHI workshops on different topics about computational UIs [33, 41, 42, 70]. This year, we extend this trajectory toward the rapidly evolving space of UI agents, multimodal interaction, and adaptive interfaces spanning both 2D and 3D settings. Discussions will highlight strategies for integrating AI into design and interaction while preserving creativity, agency, and trust, as well as enabling richer, context-aware experiences across devices and environments. The main topics of our workshop include:

- **UI Agents as Collaborative Partners in Use:** Intelligent agents are increasingly embedded in design environments and end-user applications, acting as collaborators rather than mere tools. This topic explores how UI agents can support creativity, productivity, and decision-making while preserving human oversight.
- **Multimodal Interfaces for AI-Enhanced Interaction:** With advances in speech, gesture, sketch, gaze, and haptic technologies, multimodal interaction offers new opportunities for AI-driven UIs. This topic investigates how multiple modalities can be combined to improve communication with intelligent systems, support accessibility, and create richer, more adaptive experiences.
- **Computational UIs in XR Environments:** The rise of smart AR glasses and immersive XR technologies presents unique challenges and opportunities for interface design. This topic explores how computational UIs can operate in embodied, spatial, and hybrid contexts, from adaptive overlays in augmented reality to intelligent agents situated in virtual spaces.

- **Bidirectional Alignment Between AI and Human Workflows:** Alignment in computational UIs is not one-sided: designers must adapt to AI-driven processes, while AI systems must also adapt to human practices, goals, and values. This topic examines strategies for achieving bidirectional alignment, ensuring that AI tools integrate into existing workflows while remaining transparent, trustworthy, and adaptable to human needs.

2 BACKGROUND

2.1 AI-Enhanced UI Design, Prototyping, and Workflows

The practice of designing and developing user interfaces has long benefited from systematic investigation into human cognition and interaction. Early HCI research highlighted the importance of users' mental models, affordances, and the gulfs of execution and evaluation [76, 81], which in turn guided interface design principles and development workflows. These insights informed the creation of early experimental design tools [58] and model-based UI development frameworks [19, 79], while sketching and parallel prototyping techniques were recognized for supporting creativity and mitigating design fixation [2, 11]. Collectively, these practices have shaped modern UI/UX workflows, establishing a foundation for exploring how AI can further augment design and development.

Recent advances in AI and data-driven approaches have opened new possibilities for understanding, generating, and evaluating user interfaces. Large-scale UI datasets such as Rico [7] and WebUI [102] have enabled deep learning methods for layout generation [21, 26, 48, 62, 72], code generation [71, 101], semantic modeling [1, 63, 107], and user behavior prediction [36, 38, 39, 90]. Pre-trained large multimodal models further extend these capabilities, supporting tasks that range from automated UI evaluation [13, 45, 52, 94, 100] to accessibility analysis [93] and multi-agent design workflows [18]. Prompting and fine-tuning techniques have demonstrated that general-purpose AI can adapt to UI-specific tasks, enabling richer, more adaptive interfaces while maintaining the potential for human oversight.

Alongside technical advances, researchers and practitioners are examining how AI can be effectively integrated into design and development workflows. UX professionals often envision AI as both a creative collaborator providing inspiration, design suggestions, and explorations, and as an assistant to reduce repetitive work [51]. Prototype systems and commercial tools illustrate this dual role, from AI-powered gallery interfaces for design inspiration [28, 53] to generative UI mechanisms that automate mundane tasks while jumpstarting creative processes [18, 71]. At the same time, generative UI mechanisms benefit from human-centered design; otherwise, implicit requirements might be overlooked and erroneous assumptions implicitly enacted by the models during generation [54, 109]. Emerging research also highlights opportunities for AI to facilitate collaboration between designers and developers [55, 66], though challenges remain around preserving human agency, creative control, and trust, as well as addressing biases and intellectual property concerns [51, 73].

2.2 GUI Agent

GUI agents are specialized software agents that interact with visual elements of computing environments. They are capable of executing tasks ranging from simple web navigation [108] to more advanced information-gathering operations [106], operating either within websites [9, 113], mobile interfaces [44], or across entire operating systems [103]. Contemporary research emphasizes GUI agents built on large language models (LLMs) or vision-language models (VLMs), enhanced with frameworks that handle observation processing, action execution, and control flow management [22].

The development of GUI agents has progressed rapidly. Initial efforts, including WebGPT [77] and *WIPI* [99], integrated LLMs with external information retrieval to accomplish web-related tasks. Later systems, such as WebShop [105] and Mind2Web [8], relied on direct manipulation of HTML content to perform actions online. More advanced approaches, including SeeAct [112], WebArena [114], and *Set-of-Mark* [104], utilize screenshots to extract higher-level semantic representations, enabling richer comprehension and interaction with web pages.

Despite increasing interest in GUI Agents [78, 110] and a number of products released in the area [83, 84], it remains unclear when and how these technologies will be useful for people. Recently, the HCI community has started to explore ways for humans and GUI agents to work together [23, 32], which is starting to show some advantages for working together, such as GUI agents being able to sometimes help novice users through complex user interfaces [29, 32] and especially shows promise for people who are blind where it can be otherwise difficult to get an overview of what options are available in a GUI [86]. Another challenge is in "safety" [30, 92, 98, 111], where the high-level HCI challenge can be summarized as how can GUI agents make users more efficient while not requiring users to watch and confirm each step that they make.

2.3 Adaptive User Interfaces in Mixed Reality

Prior work on adaptive user interfaces for XR has largely concentrated on how virtual elements are displayed, particularly regarding their positioning and presentation within diverse environments. There exists a broad spectrum of strategies for anchoring XR content [88], ranging from tracking-based methods [56] to marker-based techniques [31, 49, 50, 89, 91], which rely on fiducial markers or specific objects to fix virtual elements in place. Approaches grounded in geometry or semantics adopt more context-sensitive strategies by interpreting features of the surrounding scene. Geometry-driven techniques align virtual content with structural elements such as edges, planes, and meshes [12, 15, 16, 20, 25, 27, 82], whereas semantic-based strategies exploit relationships between physical entities and virtual components [3, 10, 24, 59, 65, 95] or leverage the user's current context [17], including ongoing physical activity (e.g., walking [57]) and cognitive load [67]. Another research trajectory examines how users interact with XR interfaces featuring different layout designs. For instance, Lu et al. [68, 69] and Davari et al. [5] studied the Glanceable AR paradigm, in which supplementary information is positioned at the periphery of a user's vision, remaining minimally intrusive while still accessible via quick glances. Collectively, these studies focus on determining optimal placements for predefined virtual elements within real-world contexts and arranging them adaptively. However, despite extensive work on adaptive presentations, there remains a gap in understanding which XR content should be offered. Specifically, it is still uncertain what types of virtual elements users find valuable for everyday XR tasks and how much information they prefer to engage with at a given time. This knowledge gap is partly attributable to the current immaturity of XR hardware for everyday scenarios. Nonetheless, this represents a chicken-and-egg problem: understanding potential usability can both guide and motivate the design of XR devices that better support these use cases.

Integrating generative AI models to achieve more personalized XR interfaces represents another significant research avenue. Prior studies highlight that tailoring XR experiences to individual users is crucial for ensuring both usability and meaningful engagement [64]. For example, MineXR [4] illustrates the benefits of modular application design, allowing users to assemble widget-based layouts that reflect their personal, contextual, and task-specific needs. Findings suggest that users tend to favor functionality-centric interfaces over application-centric ones, prioritizing direct access to essential features (e.g., email inboxes or recipe steps) while reducing unnecessary navigation. Similarly, research on adaptive XR interfaces [68] emphasizes context-aware adaptation, demonstrating that XR systems should dynamically respond to varying environments (e.g., office, kitchen, public spaces), activities (e.g., work, leisure, study, sports), and changing states (e.g., stationary versus mobile scenarios).

Generative AI introduces new possibilities for extending these adaptive paradigms through generative, adaptive, and mixed-initiative personalization. Unlike conventional rule-based or pre-configured adaptive systems, AI models can generate novel layouts, styles, and multimodal representations that more closely align with individual preferences. By leveraging multimodal inputs such as gaze [36, 37, 39, 75, 96, 97], gestures [74], environmental context [60], or historical interaction data [6, 61], AI-driven systems can proactively surface contextually relevant features (e.g., recipe and timer widgets while cooking or communication and scheduling tools during collaborative tasks) and adjust them in real time. This approach opens the door to XR interfaces that are not only functionally efficient but also aesthetically and emotionally aligned with user identity, preferences, and mood.

At the same time, incorporating AI models for personalized XR interfaces introduces new challenges for the HCI community. Mixed-initiative control becomes critical: while users may benefit from proactive system adaptations, excessive automation risks diminishing autonomy and transparency. Additionally, adaptive personalization may unintentionally propagate biases in training data, overfit to transient user behaviors, or pressure users into persuasive interactions. Consequently, evaluation frameworks must expand beyond traditional efficiency metrics to also consider long-term engagement, trust, and users' perceived sense of agency.

3 THE GOAL OF THE WORKSHOP

The primary objective of this workshop is to inspire the community to explore research opportunities in AI techniques and tools and their impact on design workflows and practices. Our aim is to have impact in academic, practical, and social contexts in designing and implementing high-quality, user friendly interfaces and experiences. We encourage individuals from various backgrounds, including the CHI community, adjacent academic fields, and industry practitioners to participate and think about future opportunities and challenges.

Through the discussions held during the workshop, our intention is to reflect on different ways AI will influence user interactions with interfaces and their implications on design and development requirements, tools, and practices. We draw attention to research on user behaviors, UI/UX design practices, and the latest AI capabilities to spark conversations between attendees and inspire participatory design of the preferred future state. We intend for this workshop to act as a platform that advances both new research directions and direct product impact.

4 ORGANIZERS

The organizing team has successfully organized **four** workshops at CHI on different topics of computational user interfaces [40–42, 70]. The organizing team brings a wealth of successful experience in hosting CHI workshops on related topics. Before shifting to this proposal's specific focus on human-AI-UI interactions, members of the team successfully organized workshops on computational methods for understanding, generating, and evaluating user interfaces. Leveraging this expertise, the organizers are committed to delivering a high-quality and inspiring workshop experience, ensuring smooth planning and execution.

The organizing team includes both academic and industry researchers, ensuring a balanced perspective that bridges cutting-edge research and practical applications in AI-supported design tooling. This blend of expertise allows the team to create a workshop environment that is relevant and engaging for participants from diverse backgrounds. Their collective experience will facilitate insightful discussions and foster collaborations that drive forward the research and practical application of AI-enhanced UI applications.

Kewen Peng is a Ph.D. student in the School of Computing at the University of Utah. Her research focuses on enhancing user experience in XR through human-AI-UI collaboration. She is particularly interested in how adaptive, agent-driven interfaces can support everyday tasks and enrich interaction in immersive environments. Before joining Utah, Kewen received a Bachelor's degree in Mathematics and Industrial Design from Shanghai Jiao Tong University, and conducted research at Tsinghua University and Aalto University.

Jeffery Nichols is a Research Scientist in the AI/ML group at Apple working on intelligent user interfaces. Previously he was a Staff Research Scientist at Google, working on the open-source Fuchsia operating system. His most important academic contribution recently was the creation of the RICO dataset [7]. He also worked on the PUC project [80], whose primary focus was creating a specification language that can define any device and an automatic user interface generator that can create control panels from this specification language.

Christof Lutteroth is a Reader in the Department of Computer Science at the University of Bath. His main research interests are in HCI, with a focus on immersive technology, interaction methods, and user interface design. In particular, he has a long-standing interest in methods for user interface layout. He is the director of the REal and Virtual Environments Augmentation Labs (REVEAL), the research center for immersive technology at the University of Bath.

Tiffany Knearem is an Affiliated Assistant Professor at the Mohamed bin Zayed University of Artificial Intelligence (MBZUAI) and the head of TK Research, a UX and HCI research consultancy. She holds a Ph.D. in information sciences and technology from Pennsylvania State University, advised by Prof. John M. Carroll. Her recent research interests span human-AI alignment, AI-supported design workflows, and community informatics.

Felix Kretzer is a Postdoctoral Researcher at the Karlsruhe Institute of Technology (KIT). His research focuses on leveraging generative AI to semi-automate GUI development from an end-to-end perspective. Current work spans requirements-grounded GUI generation, requirements-to-prototype mapping, and multi-agent testing of GUIs against natural-language specifications.

Jeffrey Bigham is the Philip Guo Endowed Professor at the Human-Computer Interaction Institute, Carnegie Mellon University, and also a Principal Researcher at Apple. His research spans accessibility, human-centered AI, AR/VR/XR, social computing, and interaction tools.

Alexander Maedche is Full Professor and Chair at the Karlsruhe Institute of Technology (KIT), Germany. He heads the human-centered systems lab (h-lab) that is focusing on designing human-centered systems for better work and life. He has been working on the semi-automatic design and development of intelligent systems for over 25 years, and has, among other things, shaped the concept of ontology learning. His current research areas including AI companions, agentic enterprise systems, biosignal-adaptive systems, and human-computer co-creation.

Yue Jiang is an Assistant Professor at the University of Utah. Her research focuses on computational user interface understanding, with specific interests in generating adaptive UIs for different users and contexts, AI-assisted design, and modeling human behavior.

5 PRE-WORKSHOP PLANS

Before the workshop, we will distribute a call for participation across a variety of HCI-related emailing lists and social media, like Twitter and LinkedIn. The call will invite researchers and practitioners to contribute by submitting position papers. We will also advertise the workshop at upcoming HCI conferences, among research groups, and through our professional networks. All participants are expected to submit a position paper. The submissions will be reviewed by the

Time	Session
13:30 – 13:45	Welcome and Icebreaker
13:45 – 14:25	Mini-Keynotes (4 × 7 min talks + Q&A each)
14:25 – 14:45	Impulse talks (3–4 × 5 min + Q&A each)
14:45 - 15:00	Moderated transition (introduction to brainstorming activity and card draw)
15:00 - 15:30	**Coffee Break**
15:30 – 16:00	Group brainstorming (Daily Task × Modality × Agent Role)
16:00 – 16:50	Group sharing (5 groups × 7 min presentation + 3 min Q&A)
16:50 – 17:00	Wrap-up and closing remarks

Table 1. Agenda for the workshop following the CHI 2026 format (two 90-minute blocks).

workshop organizers and committee members. The selection of participants will be based on the relevancy, innovation, and quality presented in their submissions according to workshop topics and criteria. To help candidates get familiar with the workshop's scope and goals, we have created a website https://sites.google.com/view/computational-uichi25/home, to provide information about the workshop.

6 ACCESSIBILITY

Authors whose position papers are accepted will be strongly encouraged to make their papers accessible. While they are preparing for the camera-ready version, our organizing team will help them with suggestions on how to make the documents accessible, like adding alt-texts for pictures and tables, and setting the order. To make sure the workshop is accessible to people with disabilities, we will consider adding subtitles, depending on what the participants need.

7 WORKSHOP STRUCTURE

The workshop, scheduled as a two session afternoon workshop (2 × 90 mins), will accommodate roughly 30 participants (including the organizers). The workshop will include keynotes, presentations of paper and demo submissions to the workshop, and focused group discussions on a variety of related topics.

7.1 In-Person Format

The workshop is anticipated to be in-person. Standard equipment available at the conference center will suffice for technical requirements. The workshop website https://sites.google.com/view/computational-uichi26/home, will serve as a hub for information, hosting calls for papers, program details, organizers and speakers list, and accepted papers.

7.2 Workshop Schedule

Throughout the workshop, the attendees will engage with domain experts, and the organizers will guide discussions across various domains. The tentative agenda is show in Table 1.

7.2.1 Mini-Keynotes. Instead of long keynote sessions, we will feature a series of short framing talks (around 7 minutes each plus Q&A) by organizers and invited speakers. Each talk will introduce one of the workshop's core themes—such as UI agents as collaborators, multimodal interaction, computational UIs in XR, and bidirectional alignment. This format provides participants with diverse perspectives in a concise and engaging way for interactive activities.

7.2.2 Impulse Talks from Participants. Accepted position papers will be presented as short impulse talks (around 5 minutes plus Q&A). These talks will highlight ongoing work and serve as seed ideas for the brainstorming activity. By sharing diverse perspectives from both academia and industry, the talks ensure that group activities are grounded in real-world research and practice.

7.2.3 Brainstorming Activity. Building on the themes introduced in the mini-keynotes, participants will take part in a card-based ideation game designed to explore future human–AI–UI interactions across modalities. Participants will be assigned to diverse groups based on research interests. After brief introductions and icebreaking, each group will draw cards from three pools: *daily tasks*, *interaction modalities* (e.g., 2D interfaces, speech, gesture, gaze, haptics, XR overlays), and *agent roles* (e.g., background automation, co-pilot, embodied avatar). Groups will then brainstorm and sketch interaction flows, addressing: (1) how human–AI–UI collaboration unfolds, (2) what challenges emerge (e.g., privacy, safety, trust, accessibility), and (3) what opportunities multimodality may enable.

Unlike traditional theme-based brainstorming, this activity introduces a layer of randomness and play, which helps break habitual thinking and inspire unexpected design directions.

7.2.4 Group Sharing and Wrap-Up. Each group will present their ideas (about 7 minutes plus Q&A), followed by a synthesis discussion led by the organizers to connect outcomes back to the workshop's themes. This ensures participants not only exchange perspectives but also leave with a shared sense of challenges and opportunities for future research. The playful format is designed to break habitual thinking and inspire unexpected design directions.

8 POST-WORKSHOP PLAN

After the CHI workshop, we plan to produce a report on the workshop outcome. All accepted workshop papers, presentation materials, and the synthesized Computational GUI Agent Research Agenda will be made openly available on the workshop website before and after the event, enabling the broader community to stay informed.

We also aim to pursue a post-workshop publication such as an edited volume or special journal collection (e.g., ToCHI), for which participants will be invited to submit extended versions of their contributions.

A central goal of this workshop is sustained community building among researchers and practitioners working on UI agents, multimodal interaction, and human–AI–UI integration. Following the workshop, we will establish channels for ongoing collaboration, such as a periodical email newsletter, a public GitHub repository, and a Slack or Discord community for sharing resources, discussing emerging topics, and coordinating future initiatives. These activities will be further refined together with participants during the workshop's concluding session.

9 CALL FOR PARTICIPATION

"Human-AI-UI Interactions Across Modalities" is a workshop at CHI 2026. In this two sessions afternoon workshop (2 × 90 mins), our aim is to facilitate collaboration among researchers from various sub-disciplines of HCI, bridging the gaps between HCI and adjacent fields such as ML, CV, NLP, and Software Engineering, fostering dialogue across communities and perspectives. Participation requires submitting a 4–6 page position paper in the CHI Extended Abstract format (excluding references). Submissions will be peer-reviewed by the organizers and program committee, and selected based on quality and relevance. Accepted authors will give a short impulse talk (around 5 minutes) to seed discussions and activities. At least one author of each accepted paper must register for and attend both the workshop and at least one day of the conference.

Submissions can cover but are not limited to the following topics: **(1) UI Agents as Collaborative Partners in Design and Use**; **(2) Multimodal Interfaces for AI-Enhanced Interaction**; **(3) Computational UIs in XR Environments**; **(4) Bidirectional Alignment Between AI and Human Workflows.** Accepted papers will be made accessible on the workshop website (with author consent) prior to the event. Submissions should be sent via user.interface.workshop@gmail.com.

9.1 Estimated Key Dates

- Call for participation released: December 15, 2025
- Position paper submission deadline: February 23, 2026
- Notification of acceptance: March 15, 2026
- Workshop date: April 14th, 2026

REFERENCES

[1] Chongyang Bai, Xiaoxue Zang, Ying Xu, Srinivas Sunkara, Abhinav Rastogi, Jindong Chen, et al. 2021. Uibert: Learning generic multimodal representations for ui understanding. *arXiv preprint arXiv:2107.13731* (2021).

[2] Bill Buxton. 2010. *Sketching user experiences: getting the design right and the right design.* Morgan kaufmann.

[3] Yifei Cheng, Yukang Yan, Xin Yi, Yuanchun Shi, and David Lindlbauer. 2021. Semanticadapt: Optimization-based adaptation of mixed reality layouts leveraging virtual-physical semantic connections. In *The 34th Annual ACM Symposium on User Interface Software and Technology*. 282–297.

[4] Hyunsung Cho, Yukang Yan, Kashyap Todi, Mark Parent, Missie Smith, Tanya R Jonker, Hrvoje Benko, and David Lindlbauer. 2024. Minexr: Mining personalized extended reality interfaces. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–17.

[5] Shakiba Davari, Feiyu Lu, and Doug A Bowman. 2022. Validating the benefits of glanceable and context-aware augmented reality for everyday information access tasks. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 436–444.

[6] Brendan David-John, Candace Peacock, Ting Zhang, T Scott Murdison, Hrvoje Benko, and Tanya R Jonker. 2021. Towards gaze-based prediction of the intent to interact in virtual reality. In *ACM symposium on eye tracking research and applications*. 1–7.

[7] Biplab Deka, Zifeng Huang, Chad Franzen, Joshua Hibschman, Daniel Afergan, Yang Li, Jeffrey Nichols, and Ranjitha Kumar. 2017. Rico: A Mobile App Dataset for Building Data-Driven Design Applications. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology* (Québec City, QC, Canada) *(UIST '17)*. Association for Computing Machinery, New York, NY, USA, 845–854. https://doi.org/10.1145/3126594.3126651

[8] Xiang Deng, Yu Gu, Boyuan Zheng, Shijie Chen, Samuel Stevens, Boshi Wang, Huan Sun, and Yu Su. 2023. MIND2WEB: towards a generalist agent for the web. In *Proceedings of the 37th International Conference on Neural Information Processing Systems* (New Orleans, LA, USA) *(NIPS '23)*. Curran Associates Inc., Red Hook, NY, USA.

[9] Xiang Deng, Yu Gu, Boyuan Zheng, Shijie Chen, Sam Stevens, Boshi Wang, Huan Sun, and Yu Su. 2024. Mind2web: Towards a generalist agent for the web. *NeurIPS* (2024).

[10] Zhi-Chao Dong, Wenming Wu, Zenghao Xu, Qi Sun, Guanjie Yuan, Ligang Liu, and Xiao-Ming Fu. 2021. Tailored reality: Perception-aware scene restructuring for adaptive vr navigation. *ACM Transactions on Graphics (TOG)* 40, 5 (2021), 1–15.

[11] Steven P Dow, Alana Glassco, Jonathan Kass, Melissa Schwarz, Daniel L Schwartz, and Scott R Klemmer. 2010. Parallel prototyping leads to better design results, more divergence, and increased self-efficacy. *ACM Transactions on Computer-Human Interaction (TOCHI)* 17, 4 (2010), 1–24.

[12] Ruofei Du, Eric Turner, Maksym Dzitsiuk, Luca Prasso, Ivo Duarte, Jason Dourgarian, Joao Afonso, Jose Pascoal, Josh Gladstone, Nuno Cruces, et al. 2020. DepthLab: Real-time 3D interaction with depth maps for mobile augmented reality. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. 829–843.

[13] Peitong Duan, Chin-yi Chen, Gang Li, Bjoern Hartmann, and Yang Li. 2024. UICrit: Enhancing Automated Design Evaluation with a UICritique Dataset. *arXiv preprint arXiv:2407.08850* (2024).

[14] Parvin Emami, Yue Jiang, Zixin Guo, and Luis A Leiva. 2024. Impact of Design Decisions in Scanpath Modeling. *Proceedings of the ACM on Human-Computer Interaction* 8, ETRA (2024), 1–16.

[15] Barrett Ens, Eyal Ofek, Neil Bruce, and Pourang Irani. 2015. Spatial constancy of surface-embedded layouts across multiple environments. In *Proceedings of the 3rd ACM Symposium on Spatial User Interaction*. 65–68.

[16] Andreas Fender, Philipp Herholz, Marc Alexa, and Jörg Müller. 2018. Optispace: Automated placement of interactive 3d projection mapping content. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–11.

[17] Andreas Fender, David Lindlbauer, Philipp Herholz, Marc Alexa, and Jörg Müller. 2017. Heatspace: Automatic placement of displays by empirical analysis of user behavior. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*. 611–621.

[18] Sidong Feng, Mingyue Yuan, Jieshan Chen, Zhenchang Xing, and Chunyang Chen. 2023. Designing with Language: Wireframing UI Design Intent with Generative Large Language Models. *arXiv preprint arXiv:2312.07755* (2023).

[19] Krzysztof Gajos and Daniel S. Weld. 2004. SUPPLE: Automatically Generating User Interfaces. In *Proceedings of the 9th International Conference on Intelligent User Interfaces* (Funchal, Madeira, Portugal) *(IUI '04)*. Association for Computing Machinery, New York, NY, USA, 93–100. https://doi.org/10.1145/964442.964461

[20] Ran Gal, Lior Shapira, Eyal Ofek, and Pushmeet Kohli. 2014. FLARE: Fast layout for augmented reality applications. In *2014 IEEE international symposium on mixed and augmented reality (ISMAR)*. IEEE, 207–212.

[21] Aryan Garg, Yue Jiang, and Antti Oulasvirta. 2025. Controllable gui exploration. *arXiv preprint arXiv:2502.03330* (2025).

[22] Boyu Gou, Ruohan Wang, Boyuan Zheng, Yanan Xie, Cheng Chang, Yiheng Shu, Huan Sun, and Yu Su. 2024. Navigating the digital world as humans do: Universal visual grounding for gui agents. *arXiv:2410.05243* (2024).

[23] Boyu Gou, Ruohan Wang, Boyuan Zheng, Yanan Xie, Cheng Chang, Yiheng Shu, Huan Sun, and Yu Su. 2024. Navigating the digital world as humans do: Universal visual grounding for gui agents. *arXiv preprint arXiv:2410.05243* (2024).

[24] Lei Han, Tian Zheng, Yinheng Zhu, Lan Xu, and Lu Fang. 2020. Live semantic 3d perception for immersive augmented reality. *IEEE transactions on visualization and computer graphics* 26, 5 (2020), 2012–2022.

[25] Violet Yinuo Han, Hyunsung Cho, Kiyosu Maeda, Alexandra Ion, and David Lindlbauer. 2023. Blendmr: A computational method to create ambient mixed reality interfaces. *Proceedings of the ACM on Human-Computer Interaction* 7, ISS (2023), 217–241.

[26] Lena Hegemann, Yue Jiang, Joon Gi Shin, Yi-Chi Liao, Markku Laine, and Antti Oulasvirta. 2023. Computational Assistance for User Interface Design: Smarter Generation and Evaluation of Design Ideas. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–5.

[27] Dominik Herr, Jan Reinhardt, Guido Reina, Robert Krüger, Rafael V Ferrari, and Thomas Ertl. 2018. Immersive modular factory layout planning using augmented reality. *Procedia CIRP* 72 (2018), 1112–1117.

[28] Forrest Huang, John F. Canny, and Jeffrey Nichols. 2019. Swire: Sketch-Based User Interface Retrieval. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3290605.3300334

[29] Tian Huang, Chun Yu, Weinan Shi, Zijian Peng, David Yang, Weiqi Sun, and Yuanchun Shi. 2025. Prompt2task: Automating ui tasks on smartphones from textual prompts. *ACM Transactions on Computer-Human Interaction* (2025).

[30] Xiaowei Huang, Wenjie Ruan, Wei Huang, Gaojie Jin, Yi Dong, Changshun Wu, Saddek Bensalem, Ronghui Mu, Yi Qi, Xingyu Zhao, et al. 2024. A survey of safety and trustworthiness of large language models through the lens of verification and validation. *Artificial Intelligence Review* 57, 7 (2024), 175.

[31] Ke Huo, Yuanzhi Cao, Sang Ho Yoon, Zhuangying Xu, Guiming Chen, and Karthik Ramani. 2018. Scenariot: Spatially mapping smart things within augmented reality scenes. In *Proceedings of the 2018 chi conference on human factors in computing systems*. 1–13.

[32] Faria Huq, Zora Zhiruo Wang, Frank F. Xu, Tianyue Ou, Shuyan Zhou, Jeffrey P. Bigham, and Graham Neubig. 2025. CowPilot: A Framework for Autonomous and Human-Agent Collaborative Web Navigation. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (System Demonstrations)*. Association for Computational Linguistics, 163–172. https://doi.org/10.18653/v1/2025.naacl-demo.17

[33] Yue Jiang. 2024. Computational Representations for Graphical User Interfaces. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–6.

[34] Yue Jiang. 2025. Computational representations for user interfaces. (2025).

[35] Yue Jiang, Ruofei Du, Christof Lutteroth, and Wolfgang Stuerzlinger. 2019. ORC layout: Adaptive GUI layout with OR-constraints. In *Proceedings of the 2019 CHI Conference on human factors in computing systems*. 1–12.

[36] Yue Jiang, Zixin Guo, Hamed Rezazadegan Tavakoli, Luis A Leiva, and Antti Oulasvirta. 2024. EyeFormer: predicting personalized scanpaths with transformer-guided reinforcement learning. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–15.

[37] Yue Jiang, Luis A Leiva, Paul RB Houssel, Hamed R Tavakoli, Julia Kylmälä, and Antti Oulasvirta. 2024. UEyes: an eye-tracking dataset across user interface types. *arXiv preprint arXiv:2402.05202* (2024).

[38] Yue Jiang, Luis A Leiva, Hamed Rezazadegan Tavakoli, Paul RB Houssel, Julia Kylmälä, and Antti Oulasvirta. 2023. UEyes: An Eye-Tracking Dataset across User Interface Types. In *Workshop Paper at the 2023 CHI Conference on Human Factors in Computing Systems*.

[39] Yue Jiang, Luis A Leiva, Hamed Rezazadegan Tavakoli, Paul RB Houssel, Julia Kylmälä, and Antti Oulasvirta. 2023. Ueyes: Understanding visual saliency across user interface types. In *Proceedings of the 2023 CHI conference on human factors in computing systems*. 1–21.

[40] Yue Jiang, Yuwen Lu, Tiffany Knearem, Clara E Kliman-Silver, Christof Lutteroth, Toby Jia-Jun Li, Jeffrey Nichols, and Wolfgang Stuerzlinger. 2024. Computational Methodologies for Understanding, Automating, and Evaluating User Interfaces. In *Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems (CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, Article 462, 7 pages. https://doi.org/10.1145/3613905.3636316

[41] Yue Jiang, Yuwen Lu, Christof Lutteroth, Toby Jia-Jun Li, Jeffrey Nichols, and Wolfgang Stuerzlinger. 2023. The future of computational approaches for understanding and adapting user interfaces. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–5.

[42] Yue Jiang, Yuwen Lu, Jeffrey Nichols, Wolfgang Stuerzlinger, Chun Yu, Christof Lutteroth, Yang Li, Ranjitha Kumar, and Toby Jia-Jun Li. 2022. Computational approaches for understanding, generating, and adapting user interfaces. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–6.

[43] Yue Jiang, Christof Lutteroth, Rajiv Jain, Christopher Tensmeyer, Varun Manjunatha, Wolfgang Stuerzlinger, and Vlad I Morariu. 2024. FlexDoc: Flexible Document Adaptation through Optimizing both Content and Layout. In *2024 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*. IEEE, 217–222.

[44] Yue Jiang, Eldon Schoop, Amanda Swearngin, and Jeffrey Nichols. 2023. ILuvUI: Instruction-tuned LangUage-Vision modeling of UIs from Machine Conversations. arXiv:2310.04869 [cs.HC]

[45] Yue Jiang, Eldon Schoop, Amanda Swearngin, and Jeffrey Nichols. 2025. Iluvui: Instruction-tuned language-vision modeling of uis from machine conversations. In *Proceedings of the 30th International Conference on Intelligent User Interfaces*. 861–877.

[46] Yue Jiang, Wolfgang Stuerzlinger, and Christof Lutteroth. 2021. ReverseORC: Reverse engineering of resizable user interface layouts with or-constraints. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–18.

[47] Yue Jiang, Wolfgang Stuerzlinger, Matthias Zwicker, and Christof Lutteroth. 2020. Orcsolver: An efficient solver for adaptive gui layout with or-constraints. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14.

[48] Yue Jiang, Changkong Zhou, Vikas Garg, and Antti Oulasvirta. 2024. Graph4gui: Graph neural networks for representing graphical user interfaces. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–18.

[49] Dongsik Jo and Gerard Jounghyun Kim. 2016. ARIoT: scalable augmented reality framework for interacting with Internet of Things appliances everywhere. *IEEE Transactions on Consumer Electronics* 62, 3 (2016), 334–340.

[50] Hirokazu Kato and Mark Billinghurst. 1999. Marker tracking and hmd calibration for a video-based augmented reality conferencing system. In *Proceedings 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR'99)*. IEEE, 85–94.

[51] Tiffany Knearem, Mohammed Khwaja, Yuling Gao, Frank Bentley, and Clara E Kliman-Silver. 2023. Exploring the future of design tooling: The role of artificial intelligence in tools for user experience professionals. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–6.

[52] Kristian Kolthoff, Felix Kretzer, Christian Bartelt, Alexander Maedche, and Simone Paolo Ponzetto. 2024. Interlinking User Stories and GUI Prototyping: A Semi-Automatic LLM-Based Approach. In *2024 IEEE 32nd International Requirements Engineering Conference (RE)*. 380–388. https://doi.org/10.1109/RE59067.2024.00045

[53] Kristian Kolthoff, Felix Kretzer, Christian Bartelt, Alexander Maedche, and Simone Paolo Ponzetto. 2025. GUI-ReRank: Enhancing GUI Retrieval with Multi-Modal LLM-based Reranking. arXiv:2508.03298 [cs.SE] https://arxiv.org/abs/2508.03298

[54] Kristian Kolthoff, Felix Kretzer, Christian Bartelt, Alexander Maedche, and Simone Paolo Ponzetto. 2025. GUIDE: LLM-Driven GUI Generation Decomposition for Automated Prototyping. In *2025 IEEE/ACM 47th International Conference on Software Engineering: Companion Proceedings (ICSE-Companion)*. 1–4. https://doi.org/10.1109/ICSE-Companion66252.2025.00010

[55] Felix Kretzer, Kristian Kolthoff, Christian Bartelt, Simone Paolo Ponzetto, and Alexander Maedche. 2025. Closing the Loop between User Stories and GUI Prototypes: An LLM-Based Assistant for Cross-Functional Integration in Software Development *(CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 879, 19 pages. https://doi.org/10.1145/3706598.3713932

[56] Kin Chung Kwan and Hongbo Fu. 2019. Mobi3dsketch: 3d sketching in mobile ar. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–11.

[57] Wallace S Lages and Doug A Bowman. 2019. Walking with adaptive augmented reality workspaces: design and usage patterns. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*. 356–366.

[58] James A Landay. 1996. SILK: sketching interfaces like krazy. In *Conference companion on Human factors in computing systems*. 398–399.

[59] Yining Lang, Wei Liang, and Lap-Fai Yu. 2019. Virtual agent positioning driven by scene semantics in mixed reality. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 767–775.

[60] Jong-in Lee and Wolfgang Stuerzlinger. 2025. Towards Personalized Navigation in XR: Design Recommendations to Accommodate Individual Differences. In *2025 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 241–245.

[61] Luis A Leiva. 2011. Restyling website design via touch-based interactions. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*. 599–604.

[62] Jianan Li, Jimei Yang, Aaron Hertzmann, Jianming Zhang, and Tingfa Xu. 2019. Layoutgan: Generating graphic layouts with wireframe discriminators. *arXiv preprint arXiv:1901.06767* (2019).

[63] Toby Jia-Jun Li, Lindsay Popowski, Tom Mitchell, and Brad A Myers. 2021. Screen2Vec: Semantic Embedding of GUI Screens and GUI Components. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 578, 15 pages. https://doi.org/10.1145/3411764.3445049

[64] Zhipeng Li, Christoph Gebhardt, Yves Inglin, Nicolas Steck, Paul Streli, and Christian Holz. 2024. Situationadapt: Contextual ui optimization in mixed reality with situation awareness via llm reasoning. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–13.

[65] Wei Liang, Xinzhe Yu, Rawan Alghofaili, Yining Lang, and Lap-Fai Yu. 2021. Scene-aware behavior synthesis for virtual pets in mixed reality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–12.

[66] James Lin, Tiffany Knearem, and Kristopher Giesing. 2023. Relay: A collaborative UI model for design handoff. In *Adjunct Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–3.

[67] David Lindlbauer, Anna Maria Feit, and Otmar Hilliges. 2019. Context-aware online adaptation of mixed reality interfaces. In *Proceedings of the 32nd annual ACM symposium on user interface software and technology*. 147–160.

[68] Feiyu Lu and Doug A Bowman. 2021. Evaluating the potential of glanceable ar interfaces for authentic everyday uses. In *2021 IEEE virtual reality and 3D user interfaces (VR)*. IEEE, 768–777.

[69] Feiyu Lu, Shakiba Davari, Lee Lisle, Yuan Li, and Doug A Bowman. 2020. Glanceable ar: Evaluating information access methods for head-worn augmented reality. In *2020 IEEE conference on virtual reality and 3D user interfaces (VR)*. IEEE, 930–939.

[70] Yuwen Lu, Yue Jiang, Tiffany Knearem, Clara E Kliman-Silver, Christof Lutteroth, Jeffrey Nichols, and Wolfgang Stuerzlinger. 2025. Designing and Developing User Interfaces with AI: Advancing Tools, Workflows, and Practices. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–7.

[71] Yuwen Lu, Alan Leung, Amanda Swearngin, Jeffrey Nichols, and Titus Barik. 2024. Misty: UI Prototyping Through Interactive Conceptual Blending. *arXiv preprint arXiv:2409.13900* (2024).

[72] Yuwen Lu, Ziang Tong, Qinyi Zhao, Chengzhi Zhang, and Toby Jia-Jun Li. 2023. UI Layout Generation with LLMs Guided by UI Grammar. *arXiv preprint arXiv:2310.15455* (2023).

[73] Yuwen Lu, Chengzhi Zhang, Iris Zhang, and Toby Jia-Jun Li. 2022. Bridging the Gap between UX Practitioners' work practices and AI-enabled design support tools. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–7.

[74] Mathias N Lystbæk, Thorbjørn Mikkelsen, Roland Krisztandl, Eric J Gonzalez, Mar Gonzalez-Franco, Hans Gellersen, and Ken Pfeuffer. 2024. Hands-on, Hands-off: Gaze-Assisted Bimanual 3D Interaction. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–12.

[75] Pavel Manakhov, Ludwig Sidenmark, Ken Pfeuffer, and Hans Gellersen. 2024. Gaze on the go: Effect of spatial reference frame on visual target acquisition during physical locomotion in extended reality. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–16.

[76] Brad A Myers. 1985. The importance of percent-done progress indicators for computer-human interfaces. *ACM SIGCHI Bulletin* 16, 4 (1985), 11–17.

[77] Reiichiro Nakano, Jacob Hilton, Suchir Balaji, Jeff Wu, Long Ouyang, Christina Kim, Christopher Hesse, Shantanu Jain, Vineet Kosaraju, William Saunders, Xu Jiang, Karl Cobbe, Tyna Eloundou, Gretchen Krueger, Kevin Button, Matthew Knight, Benjamin Chess, and John Schulman. 2022. WebGPT: Browser-assisted question-answering with human feedback. arXiv:2112.09332 [cs.CL] https://arxiv.org/abs/2112.09332

[78] Dang Nguyen, Jian Chen, Yu Wang, Gang Wu, Namyong Park, Zhengmian Hu, Hanjia Lyu, Junda Wu, Ryan Aponte, Yu Xia, et al. 2024. Gui agents: A survey. *arXiv preprint arXiv:2412.13501* (2024).

[79] Jeffrey Nichols, Brad A Myers, Michael Higgins, Joseph Hughes, Thomas K Harris, Roni Rosenfeld, and Mathilde Pignol. 2002. Generating remote control interfaces for complex appliances. In *Proceedings of the 15th annual ACM symposium on User interface software and technology*. 161–170.

[80] Jeffrey Nichols, Brad A. Myers, Michael Higgins, Joseph Hughes, Thomas K. Harris, Roni Rosenfeld, and Mathilde Pignol. 2002. Generating Remote Control Interfaces for Complex Appliances. In *Proceedings of the 15th Annual ACM Symposium on User Interface Software and Technology* (Paris, France) *(UIST '02)*. Association for Computing Machinery, New York, NY, USA, 161–170. https://doi.org/10.1145/571985.572008

[81] Don Norman. 2013. *The design of everyday things: Revised and expanded edition*. Basic books.

[82] Benjamin Nuernberger, Eyal Ofek, Hrvoje Benko, and Andrew D Wilson. 2016. Snaptoreality: Aligning augmented reality to the real world. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 1233–1244.

[83] OpenAI. 2025. Computer-Using Agent: Introducing a universal interface for AI to interact with the digital world. (2025). https://openai.com/index/computer-using-agent

[84] OpenAI. 2025. Introducing Operator. (2025). https://openai.com/index/introducing-operator/

[85] Yi-Hao Peng, Faria Huq, Yue Jiang, Jason Wu, Xin Yue Li, Jeffrey P Bigham, and Amy Pavel. 2024. Dreamstruct: Understanding slides and user interfaces via synthetic data generation. In *European Conference on Computer Vision*. Springer, 466–485.

[86] Yi-Hao Peng, Dingzeyu Li, Jeffrey P. Bigham, and Amy Pavel. 2025. Morae: Proactively Pausing UI Agents for User Choices. arXiv:2508.21456 [cs.HC] https://arxiv.org/abs/2508.21456

[87] Aini Putkonen, Yue Jiang, Jingchun Zeng, Olli Tammilehto, Jussi PP Jokinen, and Antti Oulasvirta. 2025. Understanding visual search in graphical user interfaces. *International Journal of Human-Computer Studies* 199 (2025), 103483.

[88] Xun Qian, Fengming He, Xiyun Hu, Tianyi Wang, Ananya Ipsita, and Karthik Ramani. 2022. Scalar: Authoring semantically adaptive augmented reality experiences in virtual reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–18.

[89] Francisco J Romero-Ramirez, Rafael Muñoz-Salinas, and Rafael Medina-Carnicer. 2018. Speeded up detection of squared fiducial markers. *Image and vision Computing* 76 (2018), 38–47.

[90] Eldon Schoop, Xin Zhou, Gang Li, Zhourong Chen, Bjoern Hartmann, and Yang Li. 2022. Predicting and Explaining Mobile UI Tappability with Vision Modeling and Saliency Analysis. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 36, 21 pages. https://doi.org/10.1145/3491102.3517497

[91] Hartmut Seichter, Julian Looser, and Mark Billinghurst. 2008. ComposAR: An intuitive tool for authoring AR applications. In *2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*. IEEE, 177–178.

[92] Constantine Stephanidis, Gavriel Salvendy, Margherita Antona, Jessie YC Chen, Jianming Dong, Vincent G Duffy, Xiaowen Fang, Cali Fidopiastis, Gino Fragomeni, Limin Paul Fu, et al. 2019. Seven HCI grand challenges. *International Journal of Human–Computer Interaction* 35, 14 (2019), 1229–1269.

[93] Maryam Taeb, Amanda Swearngin, Eldon Schoop, Ruijia Cheng, Yue Jiang, and Jeffrey Nichols. 2023. AXNav: Replaying Accessibility Tests from Natural Language. arXiv:2310.02424 [cs.HC]

[94] Maryam Taeb, Amanda Swearngin, Eldon Schoop, Ruijia Cheng, Yue Jiang, and Jeffrey Nichols. 2024. Axnav: Replaying accessibility tests from natural language. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–16.

[95] Tomu Tahara, Takashi Seno, Gaku Narita, and Tomoya Ishikawa. 2020. Retargetable AR: Context-aware augmented reality in indoor scenes based on 3D scene graph. In *2020 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, 249–255.

[96] Uta Wagner, Andreas Asferg Jacobsen, Tiare Feuchtner, Hans Gellersen, and Ken Pfeuffer. 2024. Eye-Hand Movement of Objects in Near Space Extended Reality. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–13.

[97] Yao Wang, Yue Jiang, Zhiming Hu, Constantin Ruhdorfer, Mihai Bâce, and Andreas Bulling. 2024. VisRecall++: Analysing and predicting visualisation recallability from gaze behaviour. *Proceedings of the ACM on Human-Computer Interaction* 8, ETRA (2024), 1–18.

[98] Laura Weidinger, Maribeth Rauh, Nahema Marchal, Arianna Manzini, Lisa Anne Hendricks, Juan Mateos-Garcia, Stevie Bergman, Jackie Kay, Conor Griffin, Ben Bariach, et al. 2023. Sociotechnical safety evaluation of generative ai systems. *arXiv preprint arXiv:2310.11986* (2023).

[99] Fangzhou Wu, Shutong Wu, Yulong Cao, and Chaowei Xiao. 2024. WIPI: A New Web Threat for LLM-Driven Web Agents. arXiv:2402.16965 [cs.CR] https://arxiv.org/abs/2402.16965

[100] Jason Wu, Yi-Hao Peng, Amanda Li, Amanda Swearngin, Jeffrey P Bigham, and Jeffrey Nichols. 2024. UIClip: A Data-driven Model for Assessing User Interface Design. *arXiv preprint arXiv:2404.12500* (2024).

[101] Jason Wu, Eldon Schoop, Alan Leung, Titus Barik, Jeffrey P Bigham, and Jeffrey Nichols. 2024. UICoder: Finetuning Large Language Models to Generate User Interface Code through Automated Feedback. *arXiv preprint arXiv:2406.07739* (2024).

[102] Jason Wu, Siyan Wang, Siman Shen, Yi-Hao Peng, Jeffrey Nichols, and Jeffrey P Bigham. 2023. WebUI: A Dataset for Enhancing Visual UI Understanding with Web Semantics. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–14.

[103] Tianbao Xie, Danyang Zhang, Jixuan Chen, Xiaochuan Li, Siheng Zhao, Ruisheng Cao, Toh Jing Hua, Zhoujun Cheng, Dongchan Shin, Fangyu Lei, et al. 2024. OSworld: Benchmarking multimodal agents for open-ended tasks in real computer environments. *arXiv:2404.07972* (2024).

[104] Jianwei Yang, Hao Zhang, Feng Li, Xueyan Zou, Chunyuan Li, and Jianfeng Gao. 2023. Set-of-Mark Prompting Unleashes Extraordinary Visual Grounding in GPT-4V. arXiv:2310.11441 [cs.CV] https://arxiv.org/abs/2310.11441

[105] Shunyu Yao, Howard Chen, John Yang, and Karthik Narasimhan. 2023. WebShop: Towards Scalable Real-World Web Interaction with Grounded Language Agents. arXiv:2207.01206 [cs.CL] https://arxiv.org/abs/2207.01206

[106] Ori Yoran, Samuel Joseph Amouyal, Chaitanya Malaviya, Ben Bogin, Ofir Press, and Jonathan Berant. 2024. Assistantbench: Can web agents solve realistic and time-consuming tasks? *arXiv:2407.15711* (2024).

[107] Keen You, Haotian Zhang, Eldon Schoop, Floris Weers, Amanda Swearngin, Jeffrey Nichols, Yinfei Yang, and Zhe Gan. 2024. Ferret-UI: Grounded Mobile UI Understanding with Multimodal LLMs. *arXiv preprint arXiv:2404.05719* (2024).

[108] Manzil Zaheer, Kenneth Marino, Will Grathwohl, John Schultz, Wendy Shang, Sheila Babayan, Arun Ahuja, Ishita Dasgupta, Christine Kaeser-Chen, and Rob Fergus. 2022. Learning to navigate wikipedia by taking random walks. *NeurIPS* (2022).

[109] J.D. Zamfirescu-Pereira, Eunice Jun, Michael Terry, Qian Yang, and Bjoern Hartmann. 2025. Beyond Code Generation: LLM-supported Exploration of the Program Design Space. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 153, 17 pages. https://doi.org/10.1145/3706598.3714154

[110] Chaoyun Zhang, Shilin He, Jiaxu Qian, Bowen Li, Liqun Li, Si Qin, Yu Kang, Minghua Ma, Guyue Liu, Qingwei Lin, et al. 2024. Large language model-brained gui agents: A survey. *arXiv preprint arXiv:2411.18279* (2024).

[111] Zhuohao Zhang, Eldon Schoop, Jeffrey Nichols, Anuj Mahajan, and Amanda Swearngin. 2025. 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*. 727–744.

[112] Boyuan Zheng, Boyu Gou, Jihyung Kil, Huan Sun, and Yu Su. 2024. GPT-4V(ision) is a generalist web agent, if grounded. In *Proceedings of the 41st International Conference on Machine Learning* (Vienna, Austria) *(ICML'24)*. JMLR.org.

[113] Shuyan Zhou, Frank F Xu, Hao Zhu, Xuhui Zhou, Robert Lo, Abishek Sridhar, Xianyi Cheng, Tianyue Ou, Yonatan Bisk, Daniel Fried, et al. 2024. WebArena: A Realistic Web Environment for Building Autonomous Agents. In *ICLR*.

[114] Shuyan Zhou, Frank F. Xu, Hao Zhu, Xuhui Zhou, Robert Lo, Abishek Sridhar, Xianyi Cheng, Tianyue Ou, Yonatan Bisk, Daniel Fried, Uri Alon, and Graham Neubig. 2024. WebArena: A Realistic Web Environment for Building Autonomous Agents. arXiv:2307.13854 [cs.AI] https://arxiv.org/abs/2307.13854