Computational Methodologies for Understanding, Automating, and Evaluating User Interfaces

YUE JIANG*, Aalto University, Finland
YUWEN LU*, University of Notre Dame, United States
TIFFANY KNEAREM, Google, United States
CLARA KLIMAN-SILVER, Google, United States
CHRISTOF LUTTEROTH, University of Bath, United Kingdom
TOBY JIA-JUN LI, University of Notre Dame, United States
JEFFREY NICHOLS, Apple Inc, United States
WOLFGANG STUERZLINGER, Simon Fraser University, Canada

Building on the success of the first two workshops on user interfaces (UIs) at CHI 2022 and CHI 2023, this workshop aims to advance the research field by further exploring current research trends, such as applying large language models and visual language models. Previous work has explored computational approaches to understanding and adapting UIs using constraint-based optimization models and machine learning-based data-driven approaches. In addition to further delving into these established UI research areas, we aim to trigger the exploration into the application of the latest advancements in general-purpose large language and vision-language models within the UI domain. We will encourage participants to explore novel methods for understanding, automating, and evaluating UIs. The proposed workshop seeks to bring together academic researchers and industry practitioners interested in computational approaches for UIs to discuss the needs and opportunities for future user interface algorithms, models, and applications.

ACM Reference Format:
Yue Jiang, Yuwen Lu, Tiffany Knearem, Clara 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 CHI Conference on Human Factors in Computing Systems (CHI EA '24), May 11–16, 2024, Honolulu, HI, USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3613905.3636316

1 OVERVIEW

Computational methodologies for user interfaces (UIs) have attracted significant attention in the HCI community [5, 56, 65]. Recent endeavors delve into the creation of UIs to aid design workflows [17, 28, 33, 44, 55], adaptability to varied devices and user inclinations [13, 14, 19], and techniques to decipher tasks and user actions [34, 39, 40, 51, 54]. Efforts are also being made in reverse engineering to enhance accessibility and understand UIs [10, 25, 41, 66], along with the development of innovative UI representations [19, 37].

The generation and adaptation of UIs to meet specific constraints are being pursued through optimization-based methodologies [13, 19, 25, 26, 47], while data-driven strategies are leveraging deep learning networks to enhance the

*Contributed equally.

Manuscript submitted to ACM

understanding of interface semantics on a large scale and to formulate interface designs [37, 61, 66]. As advancements continue in both these realms, UI design research is evolving into a multifaceted domain, drawing insights and contributions from specialists in optimization algorithms, data-driven modelling, software development, accessibility, and a range of UI applications. This interdisciplinary approach is garnering widespread attention and interest from the larger community.

Particularly, recent advancements in large language models (LLMs) and vision-language models (VLMs) have opened up new opportunities for computational understandings and interactions with UIs [60, 62, 63]. With impressive performance on diverse benchmarks, it is believable that these large, pre-trained models contain great potential for UI-related tasks. However, adapting these general large models to the domain of UI is still challenging. Topics like the need for up-to-date, high-quality datasets [9, 64], effective ways to pre-train a large UI model [30], or to build on top of pre-trained large models through few-shot learning [60] or finetuning still need further explorations. Such topics attract both HCI and ML researchers and contain great potential to advance both fields.

In the previous workshop at CHI 2022 [23] and CHI 2023 [22], we explored optimization-based approaches and data-driven approaches separately and focused on UI applications. In this workshop, we aim to focus on closing the gap between these two streams of approaches and fulfill the fundamental needs for future user interface research. Furthermore, although certain prior participants have investigated the prospective roles of Large Language Models (LLMs) in UI-related tasks [60, 62], we advocate for enriched dialogues encompassing both LLMs and extensive vision-language models as they can be significantly advantageous. The majority of submissions to preceding workshops predominantly focused on the technical research contributions within this realm. Nonetheless, incorporating diverse viewpoints, particularly from industry practitioners as seen in [27], can serve to diversify our dialogues and amplify the practical values for future research in this domain.

- **Language Models for UIs:** As discussed above, LLMs and VLMs contain great potential to improve performance on UI-related tasks [62]. Given LLMs' current impressive capabilities, it is also possible to create useful features with only text-based UI data (e.g. implementation code or view hierarchy). Moreover, the multi-modal nature of UI data determined its applicability to VLMs. How to collect high-quality data to pre-train or fine-tune these large models, and what are effective ways to prompt and interact with them would spark great explorations and discussions during the workshop.
- **UI Interaction Automation:** Interaction automation with UIs based on computational understandings can support a variety of useful applications [34, 36, 39, 62, 66]. With newer datasets and models available in this field [49], new potential directions emerge to improve computational performance in understanding user intents and predicting user actions on screen. For example, LLMs and VLMs have the potential to enhance automatic UI navigation.
- **UI Design and Development Tools:** Current UI design tools such as Figma[1] and Sketch[2] are mostly built on graphics rendering techniques. Computational generation and evaluation techniques have great potential to improve these tools and significantly increase practitioners' productivity. Building such tools requires a deep understanding of practitioners' processes and needs, which can serve as promising focal topics for our workshop. In addition, code generation for UIs has been a long-standing challenge. The capacity to automatically generate compilable UI code and produce appealing UI designs has the potential to revolutionize the way UIs are created and maintained.

[1]https://figma.com/
[2]https://www.sketch.com/

- **Evaluation:** Most qualitative UI evaluation methodologies, such as think aloud [11] and contextual interviews [16], do not directly fit with the data-driven nature of many AI-enabled computational tools. Furthermore, assessing the performance of LLMs and VLMs in UI-related tasks presents a challenge. It is important to develop new evaluation approaches for various UI models.
- **Human Behaviors on UIs:** Computational models simulating human behaviors on UIs, can lead to a deeper understanding of how people interact with UIs. For example, future UI models can benefit from understanding how people look at UIs [20, 21], and how people type texts on UIs to mimic human behaviors with fewer human annotations and user studies.
- **Mixed Reality User Interfaces:** Continuing our last year's workshop, with the increasing popularity of consumer mixed reality devices such as Meta Quest[3] and Vision Pro[4], computational approaches for mixed realities are becoming more popular. Although most 2D interface approaches can be directly applied to 3D, it is still challenging to understand the connection between physical objects and virtual interfaces and to further optimize the virtual interfaces to adapt to user preferences and cognitive load.

2 BACKGROUND

2.1 Large Language Models and Vision Language Models

Language models show great potential to interpret languages. Large language models (LLMs), such as GPT-4 [48] and Bard [52], can understand and follow task instructions in natural language. LLaMA [59] is an open-sourced LLM that is close to the performance of GPT-3. Other LLMs [7, 48, 58] further use machine-generated instruction-following examples to achieve better performance. The limitation of LLMs is that it cannot accept UI images as input. It is hard to explain all the elements and related info on the UI using language only. Vision-Language Models (VLMs) [1, 6, 8, 42] can bridge this gap by combining visual and textual information to understand UIs. LLaVA [42] is a VLM combining a vision model CLIP [18] and an LLM model Vicuna [7], but falls short in the generalizability to the UI domain. Spotlight [31] applies VLMs to UI tasks, including widget captioning, screen summarization, command grounding, and tappability prediction. ILuvUI [24] is a UI-focused instruction-following visual agent. It can perform many UI-related tasks, including conversations, detailed descriptions, listing available actions, predicting UI action outcomes, selecting UI elements, and goal-based planning. Our workshop will encourage participants to explore how to further advance the capabilities of LLMs and VLMs in the domain of UI understanding and interaction.

2.2 Optimization-based and Data-driven learning Approaches

Creating interfaces remains a challenge due to the growing diversity of devices and user preferences. Optimization-focused methods prioritize UI adaptation and customization to enhance the user experience across various devices and user needs. These approaches automatically generate user interfaces that adapt to device specifications and layout constraints while minimizing user effort [12, 13, 19]. Efficient layout solvers like Cassowary [2] and ORCSolver [26] enable real-time adaptive UI creation. Reverse engineering techniques play an important role in understanding and customizing existing UIs. Previous research has explored how to detect UI elements, enabling users to incorporate advanced interactions [10, 53], predicted constraints from input UI examples [4], reconstructed higher-level constraint-based specifications [45], facilitated webpage adaptation through various input modalities [54], and allowed platform- and framework-agnostic customization for graphical UIs and webpages [25]. Data-driven approaches have also been

[3]https://www.meta.com/quest/
[4]https://www.apple.com/apple-vision-pro/

employed to understand UIs for accessibility [66], , conduct UI accessibility testing [57], generate and evaluate UI design [15], perform UI retrieval [9, 17], learn design semantics [43], extract semantic representations from UIs [37], generate documents without the need for manual constraint and template definitions [28, 32], and combine programming-by-demonstration with natural language processing technologies to understand and create multimodal UIs [35, 38, 50]. Designing UIs is an iterative process, and designers seek fine-grained control over their designs while receiving constructive suggestions. Optimization-based methods offer designers more control over their UI designs, whereas data-driven approaches excel at generating diverse final results and suggestions. Our workshop aims to encourage researchers to deeply analyze the strengths and weaknesses of both approaches and explore the potential for combining them.

2.3 Datasets and Representations

The current line of research into data-driven techniques was primarily driven by the mobile UI dataset RICO [9], consisting of over 72,000 distinct UI screens. This dataset has been explored for data-driven applications in various aspects of UI design, including layout generation, code generation, UI modeling, and perception prediction. Subsequently, several datasets have emerged, most of which are either expansions or refined versions of RICO [3, 29]. Recently, Moran et al. introduced a fresh dataset named CLARITY [46], which focuses on functional UI descriptions. Current development of VLMs/LLMs require more efforts on new datasets to broaden the scope of this field. Having an effective representation of UIs is important for understanding and generating UIs. For instance, UIBERT [3] has developed generic feature representations for UI elements, while Screen2Vec [37] has created semantic representations for UIs. On the other hand, ORC Layout [19] represents UIs using an extended constraint system with OR-constraints. There is potential for further research to explore new UI representations to advance the capabilities of VLMs and LLMs in UI design.

3 THE GOAL OF THE WORKSHOP

The primary objective of this workshop is to inspire the community to explore potential research opportunities in UI-related topics. Our aim is to have impact in academic, practical, and social contexts in the UI domain. We want to encourage individuals from various backgrounds, including the CHI community, neighboring academic fields, and industry professionals to participant and think about future directions on this topic. Through the discussions held during the workshop, our intention is to draw attention to relevant work that sits at the intersection of HCI, other related disciplines such as Machine Learning and Software Engineering, and the expertise of industry practitioners. We aspire for this workshop to act as a platform that fosters our community's growth.

4 ORGANIZERS

Yue Jiang is a Ph.D. student in Intelligent Systems supervised by Prof. Antti Oulasvirta at Aalto University, Finland. Her main research interests are in adaptive user interfaces, AI-assisted design and eye tracking. Her recent work with Prof. Wolfgang Stuerzlinger and Prof. Christof Lutteroth focuses on adaptive GUI layouts based on OR-Constraints (ORC).
Yuwen Lu is a Ph.D. student in the Department of Computer Science and Engineering at the University of Notre Dame, working on using data-driven approaches for understanding and generating user interfaces to support UX research and design work. Before joining Notre Dame, Yuwen received a Master's degree in Human-Computer Interaction from Carnegie Mellon University.

Tiffany Knearem is a User Experience Researcher on the Material Design team at Google, where she investigates designer-developer collaboration, design tooling and opportunities for AI in the design space. She received her PhD in Information Sciences and Technology from Pennsylvania State University in 2021, with a focus on Human-Computer Interaction. There, she specialized in topics related to community informatics and creativity support. Her undergraduate education is in the humanities, and she received dual Bachelor of Arts degrees in Psychology and East Asian Languages and Cultures from the University of Kansas.

Clara Kliman-Silver is a Staff UX Researcher at Google who studies design teams, design systems, UX tools, and designer-developer collaboration. She specializes in participatory design and generative methods to investigate workflows, understand designer-developer experiences, and imagine ways to create UIs. In previous roles, she has conducted research on developer tools, artificial intelligence, and healthcare. Clara holds a Bachelors of Science in Cognitive Science from Brown University.

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.

Toby Jia-Jun Li is an Assistant Professor in the Department of Computer Science and Engineering at the University of Notre Dame and the Director of the SaNDwich Lab. Toby and his group use human-centered methods to design, build, and study human-AI collaborative systems. In the domain of this workshop, Toby has recently done work in building interactive task learning agents that learn from the user's demonstrations on GUIs and natural language instructions about GUIs [34, 38], graphs models for representing and grounding natural language instructions about GUIs [35], and semantic embedding techniques for modeling GUIs [37].

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 [9]. He also worked on the PUC project [47], 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.

Wolfgang Stuerzlinger is a Professor at the School of Interactive Arts + Technology at Simon Fraser University. His work aims to gain a deeper understanding of and to find innovative solutions for real-world problems. Current research projects include better 3D interaction techniques for Virtual and Augmented Reality applications, new human-in-the-loop systems for big data analysis, the characterization of the effects of technology limitations on human performance, investigations of human behaviors with occasionally failing technologies, user interfaces for versions, scenarios, and alternatives, and new Virtual/Augmented Reality hardware and software.

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 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

Time	Session
9:00 - 9:30	Introduction of workshop organizers, participants, topics, and goals
9:30 - 10:30	Keynote 1 by an invited speaker
10:30 - 11:00	Coffee break
11:00 - 12:00	Paper Presentation
12:00 - 12:30	Group discussion
12:30 - 13:30	Lunch
13:30 - 14:30	Keynote 2 by an invited speaker
14:30 - 15:30	Paper Presentation
15:30 - 16:00	Coffee break
16:00 - 17:00	Group discussion
17:00 - 17:30	Discussion group report back, wrap-up
17:30	Dinner (optional)

Table 1. Tentative agenda of the workshop

with the workshop's scope and goals, we have created a website https://sites.google.com/view/computational-uichi24, 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 for one day, will accommodate roughly 30 participants (including the organizers). The workshop will include two keynotes, presentations of workshop papers, and focused discussions on various topics.

7.1 Hybrid Format and Asynchronous Engagement

The workshop is anticipated to adopt a *hybrid* format, welcoming the majority of the participants in person. For those who cannot attend physically, provisions for synchronous remote involvement will be in place. All sessions will be broadcast live, with dedicated virtual "breakout rooms" for discussions involving remote attendees. Standard equipment available at the conference center will suffice for technical requirements. The workshop website https://sites.google.com/view/computational-uichi24, will serve as a hub for synchronous engagement, hosting calls for papers, program details, organizers and speakers list, and pre-prints of 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 Keynotes. We will invite two keynote speakers who are experts currently working on UI-related topics. Each will give a talk for 30 minutes, followed by an extensive Q&A and interactive discussion.

7.2.2 Paper Presentations. Accepted papers will be categorized based on their themes for presentation. We will select the best two position papers in each category for a full presentation, each allotted a 10-minute slot. Other selected papers will have lightning talks with 1-minute slots. The duration and number of presentations will be adjusted as needed to accommodate the number of accepted submissions and accommodate the technical requirements of remote presenters.

7.2.3 Breakout Group Discussions. After each presentation section, participants will be divided into smaller discussion groups. The groups will be divided differently after each session to help participants get to communicate with more people. Participants can also suggest and create new groups based on their interests.

8 POST-WORKSHOP PLAN

After the CHI workshop, we plan to produce a report on the workshop outcome. The workshop papers and results will be available on the website before and after the workshop, providing opportunities for a larger audience to get updated on the events and results of our workshop. We may seek opportunities for an edited book or a special issue in a selected journal, *e.g.*, ToCHI, where the participants will be encouraged to publish their work.

A central goal of this workshop is community building for researchers and practitioners in this area. After the workshop, we plan to create a platform for community members to continue the discussion and share resources. Potential options may include a periodical email newsletter, a public GitHub repository, or a Slack/Discord channel. Participants and organizers will discuss the next steps at the workshop.

9 CALL FOR PARTICIPATION

"Computational Methodologies for Understanding, Automating, and Evaluating User Interfaces" is a workshop at CHI 2024. In this one-day workshop, 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 SE. We welcome participants working on algorithm and model development or application creation, and we encourage engagement from both industry and academia. Our primary goal is to encourage discussions regarding the future potential and requirements of computational approaches for user interfaces.

We invite researchers and practitioners to contribute by submitting a 4–6 page position paper in the double-column CHI Extended Abstract format (excluding references) to participate in the workshop. We will have a peer-review process, with each submission reviewed by at least two committee members or organizers. Selection for submission will be based on the criteria of quality and relevance. Participants should follow the instruction on the website and submit the position papers via user.interface.workshop@gmail.com. Submissions can cover but are not limited to the following topics:

- **Language Models for UIs** We invite explorations into leveraging LLMs and VLMs for enhancing UI-related tasks. Contributions can focus on deriving impactful features using text-based UI data, exploring multi-modal UI data, and discussions surrounding high-quality data acquisition and optimal interaction methods with these models.
- **UI Interaction Automation** Position papers are welcome on interaction automation enhancing a variety of applications, delving into new directions provided by emerging datasets and models. Contributions could consider how the integration of LLMs and VLMs can revolutionize automatic UI navigation.
- **UI Design and Development Tools** We call for insights into how computational generation and evaluation techniques can refine existing design tools and boost practitioners' productivity. We are particularly interested in

submissions addressing the challenges of automatic UI code generation and offering a profound understanding of user needs and processes.

- **UI Evaluation** There is a pressing need for innovative qualitative UI evaluation methodologies that align with AI-enabled computational tools. We encourage submissions that offer novel evaluation methodologies to assess LLMs and VLMs performance in UI-related tasks.
- **Human Behaviors on UIs** We welcome studies and models that simulate human behaviors on UIs. Such contributions will yield deeper insights into user interactions and can guide the development of future UI models that mimic human behaviors with minimal reliance on annotations and user studies.
- **Mixed Reality User Interfaces** With the rising prominence of mixed reality devices, we invite discussions and contributions focusing on computational approaches for mixed realities. Emphasis on understanding and optimizing connections between physical and virtual interfaces and adapting to user preferences and cognitive load in 3D interfaces would be particularly valued.

We will recruit researchers and practitioners in this field as program committee members to review submissions together. We will select submissions based on quality, novelty, and topic fit while aiming for a balance of different perspectives. Accepted papers will optionally be available on the workshop website (with the author's consent). At least one author of each accepted position paper must register and attend the workshop and register for at least one day of the conference. The workshop will use a hybrid structure. We will broadcast the workshop live for remote participants and make the recordings available on the website after the workshop. The authors of each accepted position paper will have about 8 minutes for a live (or pre-recorded) presentation of their work followed by an additional 2-minute Q&A.

9.1 Estimated Key Dates

- Call for participation released: December 15, 2023
- Position paper submission deadline: February 23, 2024
- Notification of acceptance: March 15, 2024
- Workshop date: April 23 or April 28, 2024

REFERENCES

[1] Jean-Baptiste Alayrac, Jeff Donahue, Pauline Luc, Antoine Miech, Iain Barr, Yana Hasson, Karel Lenc, Arthur Mensch, Katie Millican, Malcolm Reynolds, Roman Ring, Eliza Rutherford, Serkan Cabi, Tengda Han, Zhitao Gong, Sina Samangooei, Marianne Monteiro, Jacob Menick, Sebastian Borgeaud, Andrew Brock, Aida Nematzadeh, Sahand Sharifzadeh, Mikolaj Binkowski, Ricardo Barreira, Oriol Vinyals, Andrew Zisserman, and Karen Simonyan. 2022. Flamingo: a Visual Language Model for Few-Shot Learning. arXiv:2204.14198 [cs.CV]

[2] Greg J. Badros, Alan Borning, and Peter J. Stuckey. 2001. The Cassowary Linear Arithmetic Constraint Solving Algorithm. *ACM Trans. Comput.-Hum. Interact* 8, 4 (2001), 267–306. https://doi.org/10.1145/504704.504705

[3] 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).

[4] Pavol Bielik, Marc Fischer, and Martin Vechev. 2018. Robust Relational Layout Synthesis from Examples for Android. *Proc. ACM Program. Lang.* 2, OOPSLA, Article 156 (Oct. 2018), 29 pages. https://doi.org/10.1145/3276526

[5] Alan Borning and Robert Duisberg. 1986. Constraint-Based Tools for Building User Interfaces. *ACM Trans. Graph.* 5, 4 (Oct. 1986), 345–374. https://doi.org/10.1145/27623.29354

[6] Xi Chen, Xiao Wang, Soravit Changpinyo, AJ Piergiovanni, Piotr Padlewski, Daniel Salz, Sebastian Goodman, Adam Grycner, Basil Mustafa, Lucas Beyer, Alexander Kolesnikov, Joan Puigcerver, Nan Ding, Keran Rong, Hassan Akbari, Gaurav Mishra, Linting Xue, Ashish Thapliyal, James Bradbury, Weicheng Kuo, Mojtaba Seyedhosseini, Chao Jia, Burcu Karagol Ayan, Carlos Riquelme, Andreas Steiner, Anelia Angelova, Xiaohua Zhai, Neil Houlsby, and Radu Soricut. 2023. PaLI: A Jointly-Scaled Multilingual Language-Image Model. arXiv:2209.06794 [cs.CV]

[7] Wei-Lin Chiang, Zhuohan Li, Zi Lin, Ying Sheng, Zhanghao Wu, Hao Zhang, Lianmin Zheng, Siyuan Zhuang, Yonghao Zhuang, Joseph E Gonzalez, et al. 2023. Vicuna: An open-source chatbot impressing gpt-4 with 90%* chatgpt quality. *See https://vicuna. lmsys. org (accessed 14 April 2023)* (2023).

[8] Wenliang Dai, Junnan Li, Dongxu Li, Anthony Meng Huat Tiong, Junqi Zhao, Weisheng Wang, Boyang Li, Pascale Fung, and Steven Hoi. 2023. InstructBLIP: Towards General-purpose Vision-Language Models with Instruction Tuning. arXiv:2305.06500 [cs.CV]

[9] 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

[10] Morgan Dixon and James Fogarty. 2010. Prefab: Implementing Advanced Behaviors Using Pixel-Based Reverse Engineering of Interface Structure. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '10)*. Association for Computing Machinery, New York, NY, USA, 1525–1534. https://doi.org/10.1145/1753326.1753554

[11] Mingming Fan, Serina Shi, and Khai N Truong. 2020. Practices and Challenges of Using Think-Aloud Protocols in Industry: An International Survey. *Journal of Usability Studies* 15, 2 (2020).

[12] Krzysztof Gajos and Daniel Weld. 2005. Preference Elicitation for Interface Optimization. *UIST: Proceedings of the Annual ACM Symposium on User Interface Softaware and Technology*, 173–182. https://doi.org/10.1145/1095034.1095063

[13] 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

[14] Krzysztof Z. Gajos, Daniel S. Weld, and Jacob O. Wobbrock. 2010. Automatically Generating Personalized User Interfaces With Supple, In Proceedings of the 9th International Conference on Intelligent User Interfaces. *Artif. Intell* 174, 12-13, 910–950. https://doi.org/10.1016/j.artint.2010.05.005

[15] 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* (<conf-loc>, <city>Hamburg</city>, <country>Germany</country>, </conf-loc>) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 430, 5 pages. https://doi.org/10.1145/3544549.3583960

[16] Ben Heuwing, Thomas Mandl, and Christa Womser-Hacker. 2016. Combining contextual interviews and participative design to define requirements for text analysis of historical media. (2016).

[17] 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

[18] Gabriel Ilharco, Mitchell Wortsman, Ross Wightman, Cade Gordon, Nicholas Carlini, Rohan Taori, Achal Dave, Vaishaal Shankar, Hongseok Namkoong, John Miller, Hannaneh Hajishirzi, Ali Farhadi, and Ludwig Schmidt. 2021. *OpenCLIP*. https://doi.org/10.5281/zenodo.5143773 If you use this software, please cite it as below.

[19] 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* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300643

[20] Yue Jiang, Luis A. Leiva, Hamed Rezazadegan Tavakoli, Paul R. B. 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* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 285, 21 pages. https://doi.org/10.1145/3544548.3581096

[21] 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. (2023).

[22] 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* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 367, 5 pages. https://doi.org/10.1145/3544549.3573805

[23] 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 *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, Article 74, 6 pages. https://doi.org/10.1145/3491101.3504030

[24] 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]

[25] 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* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 316, 18 pages. https://doi.org/10.1145/3411764.3445043

[26] 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* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376610

[27] 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.

[28] Hsin-Ying Lee, Lu Jiang, Irfan Essa, Phuong B. Le, Haifeng Gong, Ming-Hsuan Yang, and Weilong Yang. 2020. Neural Design Network: Graphic Layout Generation with Constraints. In *Computer Vision – ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III* (Glasgow, United Kingdom). Springer-Verlag, Berlin, Heidelberg, 491–506. https://doi.org/10.1007/978-3-030-58580-8_29

[29] Luis A. Leiva, Asutosh Hota, and Antti Oulasvirta. 2020. Enrico: A High-quality Dataset for Topic Modeling of Mobile UI Designs. In *Proc. MobileHCI Adjunct.*

[30] Gang Li and Yang Li. 2022. Spotlight: Mobile UI Understanding using Vision-Language Models with a Focus. *arXiv preprint arXiv:2209.14927* (2022).

[31] Gang Li and Yang Li. 2022. Spotlight: Mobile UI Understanding using Vision-Language Models with a Focus. *ArXiv* abs/2209.14927 (2022). https://api.semanticscholar.org/CorpusID:252595735

[32] 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).

[33] Jianan Li, Jimei Yang, Jianming Zhang, Chang Liu, Christina Wang, and Tingfa Xu. 2021. Attribute-Conditioned Layout GAN for Automatic Graphic Design. *IEEE Transactions on Visualization and Computer Graphics* 27, 10 (oct 2021), 4039–4048. https://doi.org/10.1109/TVCG.2020.2999335

[34] Toby Jia-Jun Li, Amos Azaria, and Brad A. Myers. 2017. SUGILITE: Creating Multimodal Smartphone Automation by Demonstration. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 6038–6049. https://doi.org/10.1145/3025453.3025483

[35] Toby Jia-Jun Li, Igor Labutov, Xiaohan Nancy Li, Xiaoyi Zhang, Wenze Shi, Wanling Ding, Tom M. Mitchell, and Brad A. Myers. 2018. APPINITE: A Multi-Modal Interface for Specifying Data Descriptions in Programming by Demonstration Using Natural Language Instructions. In *2018 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*. 105–114. https://doi.org/10.1109/VLHCC.2018.8506506

[36] Toby Jia-Jun Li, Tom Mitchell, and Brad Myers. 2020. Interactive Task Learning from GUI-Grounded Natural Language Instructions and Demonstrations. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. Association for Computational Linguistics, Online, 215–223. https://doi.org/10.18653/v1/2020.acl-demos.25

[37] 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

[38] Toby Jia-Jun Li, Marissa Radensky, Justin Jia, Kirielle Singarajah, Tom M. Mitchell, and Brad A. Myers. 2019. PUMICE: A Multi-Modal Agent That Learns Concepts and Conditionals from Natural Language and Demonstrations. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 577–589. https://doi.org/10.1145/3332165.3347899

[39] Toby Jia-Jun Li and Oriana Riva. 2018. KITE: Building conversational bots from mobile apps. In *Proceedings of the 16th ACM International Conference on Mobile Systems, Applications, and Services (MobiSys 2018)*. ACM.

[40] Yang Li, Jiacong He, Xin Zhou, Yuan Zhang, and Jason Baldridge. 2020. Mapping Natural Language Instructions to Mobile UI Action Sequences. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. ACL, Online, 8198–8210. https://doi.org/10.18653/v1/2020.acl-main.729

[41] Yang Li, Gang Li, Luheng He, Jingjie Zheng, Hong Li, and Zhiwei Guan. 2020. Widget Captioning: Generating Natural Language Description for Mobile User Interface Elements. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. ACL, Online, 5495–5510. https://doi.org/10.18653/v1/2020.emnlp-main.443

[42] Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. 2023. Visual instruction tuning. *arXiv preprint arXiv:2304.08485* (2023).

[43] Thomas F Liu, Mark Craft, Jason Situ, Ersin Yumer, Radomir Mech, and Ranjitha Kumar. 2018. Learning design semantics for mobile apps. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*. 569–579.

[44] 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.

[45] Christof Lutteroth. 2008. Automated Reverse Engineering of Hard-Coded GUI Layouts. In *Proceedings of the Ninth Conference on Australasian User Interface - Volume 76* (Wollongong, Australia) *(AUIC '08)*. Australian Computer Society, Inc., AUS, 65–73. https://doi.org/10.5555/1378337.1378350

[46] Kevin Moran, Ali Yachnes, George Purnell, Junayed Mahmud, Michele Tufano, Carlos Bernal Cardenas, Denys Poshyvanyk, and Zach H'Doubler. 2022. An Empirical Investigation into the Use of Image Captioning for Automated Software Documentation. In *2022 IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER)*. 514–525. https://doi.org/10.1109/SANER53432.2022.00069

[47] 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

[48] OpenAI. 2023. GPT-4 Technical Report. *arXiv preprint arXiv:2303.08774* (2023).

[49] Christopher Rawles, Alice Li, Daniel Rodriguez, Oriana Riva, and Timothy Lillicrap. 2023. Android in the Wild: A Large-Scale Dataset for Android Device Control. *arXiv preprint arXiv:2307.10088* (2023).

[50] Ritam Jyoti Sarmah, Yunpeng Ding, Di Wang, Cheuk Yin Phipson Lee, Toby Jia-Jun Li, and Xiang 'Anthony' Chen. 2020. Geno: A Developer Tool for Authoring Multimodal Interaction on Existing Web Applications. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '20)*. Association for Computing Machinery, New York, NY, USA, 1169–1181. https://doi.org/10.1145/3379337.3415848

[51] Alborz Rezazadeh Sereshkeh, Gary Leung, Krish Perumal, Caleb Phillips, Minfan Zhang, Afsaneh Fazly, and Iqbal Mohomed. 2020. VASTA: A Vision and Language-Assisted Smartphone Task Automation System. In *Proceedings of the 25th International Conference on Intelligent User Interfaces* (Cagliari, Italy) *(IUI '20)*. Association for Computing Machinery, New York, NY, USA, 22–32. https://doi.org/10.1145/3377325.3377515

[52] Y Societ. [n.d.]. What Bard is. https://ai.google/static/documents/google-about-bard.pdf. Accessed: 2023-10-10.

[53] Wolfgang Stuerzlinger, Olivier Chapuis, Dusty Phillips, and Nicolas Roussel. 2006. User Interface Façades: Towards Fully Adaptable User Interfaces. In *Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology* (Montreux, Switzerland) *(UIST '06)*. Association for Computing Machinery, New York, NY, USA, 309–318. https://doi.org/10.1145/1166253.1166301

[54] Amanda Swearngin, Amy J. Ko, and James Fogarty. 2017. Genie: Input Retargeting on the Web through Command Reverse Engineering. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 4703–4714. https://doi.org/10.1145/3025453.3025506

[55] Amanda Swearngin, Chenglong Wang, Alannah Oleson, James Fogarty, and Amy J. Ko. 2020. *Scout: Rapid Exploration of Interface Layout Alternatives through High-Level Design Constraints*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376593

[56] Pedro Szekely, Ping Luo, and Robert Neches. 1993. Beyond Interface Builders: Model-Based Interface Tools. In *Proceedings of the INTERCHI '93 Conference on Human Factors in Computing Systems* (Amsterdam, The Netherlands) *(INTERCHI '93)*. IOS Press, NLD, 383–390.

[57] 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]

[58] Rohan Taori, Ishaan Gulrajani, Tianyi Zhang, Yann Dubois, Xuechen Li, Carlos Guestrin, Percy Liang, and Tatsunori B. Hashimoto. 2023. Stanford Alpaca: An Instruction-following LLaMA model. https://github.com/tatsu-lab/stanford_alpaca.

[59] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. 2023. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971* (2023).

[60] Bryan Wang, Gang Li, and Yang Li. 2023. Enabling conversational interaction with mobile ui using large language models. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–17.

[61] Bryan Wang, Gang Li, Xin Zhou, Zhourong Chen, Tovi Grossman, and Yang Li. 2021. Screen2Words: Automatic Mobile UI Summarization with Multimodal Learning. In *The 34th Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '21)*. Association for Computing Machinery, New York, NY, USA, 498–510. https://doi.org/10.1145/3472749.3474765

[62] Hao Wen, Yuanchun Li, Guohong Liu, Shanhui Zhao, Tao Yu, Toby Jia-Jun Li, Shiqi Jiang, Yunhao Liu, Yaqin Zhang, and Yunxin Liu. 2023. Empowering LLM to use Smartphone for Intelligent Task Automation. *arXiv preprint arXiv:2308.15272* (2023).

[63] Hao Wen, Hongming Wang, Jiaxuan Liu, and Yuanchun Li. 2023. DroidBot-GPT: GPT-powered UI Automation for Android. *arXiv preprint arXiv:2304.07061* (2023).

[64] 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.

[65] Brad Vander Zanden and Brad A. Myers. 1990. Automatic, Look-and-Feel Independent Dialog Creation for Graphical User Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Seattle, Washington, USA) *(CHI '90)*. Association for Computing Machinery, New York, NY, USA, 27–34. https://doi.org/10.1145/97243.97248

[66] Xiaoyi Zhang, Lilian de Greef, Amanda Swearngin, Samuel White, Kyle Murray, Lisa Yu, Qi Shan, Jeffrey Nichols, Jason Wu, Chris Fleizach, et al. 2021. Screen recognition: Creating accessibility metadata for mobile applications from pixels. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–15.