Research Statement Yue Jiang

Improving Human Capabilities with Intelligent User Interfaces

Traditional user interfaces (UIs) often follow a "one-size-fits-all" design, failing to accommodate the diverse needs and contexts of all their users. What if interfaces could instead dynamically understand and adapt to individual users, enhancing their capabilities across a wide range of tasks and contexts? As the diversity of user needs expands, the challenge of designing UIs that accommodate varying contexts becomes increasingly complex. Data-driven AI methods may offer a way to design UIs that align with users' goals. Current AI methods, however, often fall short of capturing the full complexity of human needs, particularly when considering domain-specific knowledge and user-specific requirements.

Drawing on my expertise in human-computer interaction, computer vision, and deep learning, I aim **to develop human-centered technologies that enhance human capabilities by adapting to users and their contexts.** By integrating computational models with domain knowledge, I aim to use AI models to enhance human capabilities while maintaining user control. These computational approaches enable UIs to dynamically adapt to users and their contexts, promoting effective human-AI collaboration.

I have published my work in **top-tier Human-Computer Interaction (HCI) venues such as CHI, UIST, and CSCW, leading Computer Vision (CV) conferences like CVPR and ECCV, and a Programming Tool conference VL/HCC, including a CHI paper with best paper honorable mention and a CVPR oral paper (top 5%).** I am highly collaborative and have many long-term collaborators with publications from academic institutions (e.g., Carnegie Mellon University, Simon Fraser University, Max Planck Institute) and industry (e.g., Apple, Adobe Research, Google).

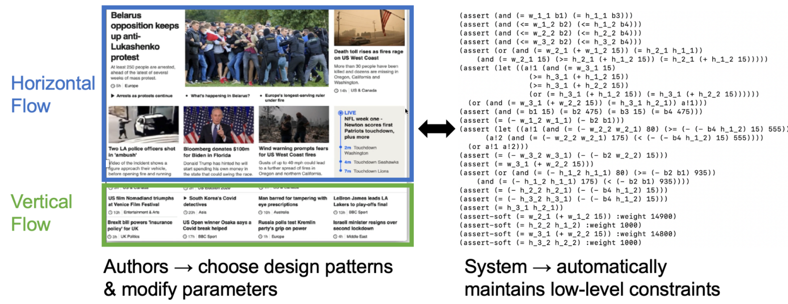
My research centers on the development of intelligent UIs in two primary areas. First, I focus on creating **computational representations** that capture the essential properties of UI design. These representations minimize the need for manual specification and enhance model performance in downstream applications and adaptive systems. They a) enable UIs to dynamically adapt to users and their contexts [2, 3, 4, 5, 10, 22], and b) integrate domain-specific knowledge into AI models [8, 12, 16, 23]. This integration allows for expert guidance while ensuring that users retain control over their interactions.

Moreover, I develop models that **simulate and predict human behaviors to facilitate automatic personalized adaptation**. These models encompass a variety of human behaviors, including simulating eye-tracking patterns [6, 9, 13, 19, 20, 21], reconstructing humans and objects from images, videos, or sensing signals [1, 7, 18], and creating interactive agents that mimic user interactions and/or respond to user requirements [11, 14, 15, 17]. By simulating these behaviors, we can achieve personalized optimization and enhance the AI's ability to adapt and respond to user needs.

COMPUTATIONAL REPRESENTATION

Advancing the field of UIs necessitates a shift from generic data-driven approaches to representations that embed domain-specific knowledge. A key challenge is **capturing the essential properties and characteristics of UI designs without requiring manual encoding**. Meanwhile, these representations must **effectively support operations needed for downstream applications**, such as learning, inference, optimization, and adaptation of UIs. Computational representation forms the backbone of how we understand, manipulate, and optimize UIs, as different representations capture unique explanatory factors and data characteristics. Structural properties—such as alignment, grouping, and layout constraints—are challenging for current deep learning methods to capture. Effective representations can capture human insight, leading to the development of improved models to assist humans. Computational methods can use such computational representations to identify and disentangle underlying explanatory factors, ultimately enhancing the computation of UIs.

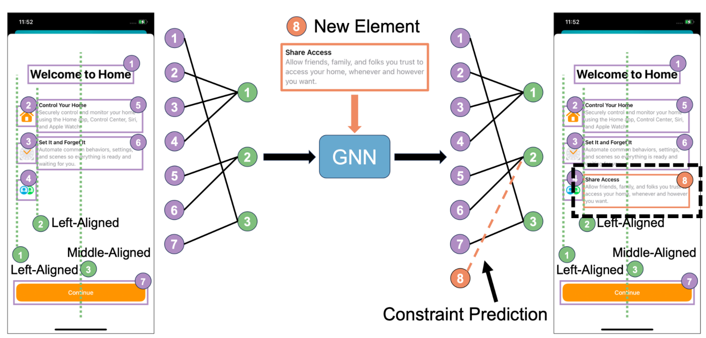
Constraint-Based Optimization for Adaptation to Contexts and User Preferences. As user needs continue diversifying, designing adaptive UIs across different contexts becomes increasingly complex. Through an elicitation study, I discovered that designers often create multiple layouts manually for various devices and user needs, a tedious and errorprone process [10]. Despite advanced design tools, the complexity of managing numerous parameters and layout models leads designers to favor manual adjustments, exacerbating their workload while increasing the risk of errors. To address this challenge, In my early work, I developed the OR-Constraint (ORC) Layout representation that **automates UI adaptation, preserving design consistency without requiring intervention from a UI designer or programmer** [2]. ORC Layout introduces disjunctive constraints, enabling complex adaptive design patterns where UI elements can automatically adjust to various screen sizes. It also accommodates user preferences by integrating automatic content generation, resulting in flexible layouts that cater to both authors and end-users [5]. This approach significantly reduces the need for manual synchronization across different versions, minimizing errors while allowing designers to focus on high-level design patterns instead of low-level constraints. However, constraint-based optimization methods typically require extensive manual specifications for initial designs and cannot offer actionable suggestions to users.



I developed a constraint-based domain-specific language that enables users to simply select patterns, while our system automatically generates and maintains the underlying low-level constraints.

Graph Neural Networks for Human-in-the-Loop AI Assistance. While current deep learning methods can generate UI designs directly from data, they limit designers' control and present challenges in modifying the generated UIs. Therefore, my objective is to create a representation that can **maximally exploit domain knowledge, particularly using stable, universal UI design characteristics without learning from scratch while simultaneously contingent design tendencies without manual specification**.

I have studied how to represent universal principles of layouts as constraints in a graph, thus reducing the need to learn them from scratch [8]. In the graph, UI elements are represented as nodes linked to constraint nodes that capture essential design principles, such as alignment and grouping. This representation incorporates both UI element properties and layout constraints. At the same time, genre-specific tendencies are learned by applying graph neural networks (GNNs) to capture the design features unique to UIs, improving the understanding of relationships between UI elements and the layout. This approach bridges the gap between AI capabilities and domain knowledge, enabling a more iterative and collaborative design process. It enhances the UI design process by maintaining designers' creative freedom while allowing them to benefit from AI-driven feedback and efficiency. The approach supports UI autocompletion by iteratively predicting the size and position of unplaced elements and providing confidence levels to help designers prioritize element placement. This iterative process allows designers to adjust AI suggestions in real time, with updates reflected in the graph representation, leading to more refined results.



I proposed a graph-based representation that links UI element properties with constraints (e.g., alignment) to capture the visual, spatial, and semantic structure of a UI. This bipartite graph comprises element nodes (purple) representing UI properties and constraint nodes (green) that can be processed by graph neural networks (GNNs) to predict constraints for new elements.

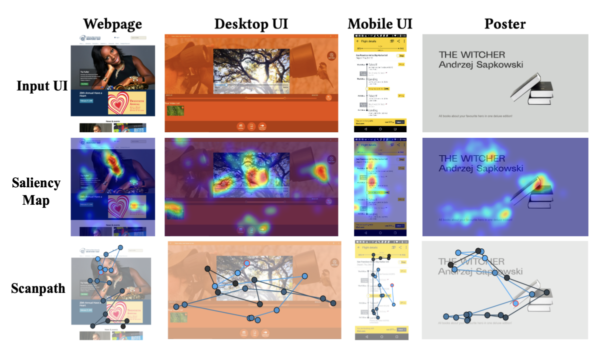
Diffusion Model for Controllable UI Exploration. Recent advances in generative AI models offer new opportunities for creative control in design generation but often struggle with imprecise inputs. During the early stages of interface design, designers need to produce multiple sketches to explore a design space. Existing design tools, however, often fail to support this critical stage, because they insist on specifying more details than necessary. However, recent advances in generative AI in practice fail because expressing loose ideas in a prompt is impractical. To bridge this gap, I proposed a diffusion-based approach for generating interface sketches with minimal effort [12]. My key insight is that we need to have a neural representation that can **take flexible inputs and produce diverse outputs**. Thus, I integrate a diffusion model with two distinct adapters—modular mechanisms that provide targeted guidance during image generation. One adapter controls local properties, such as the positioning and type of GUI elements, while the other governs global properties, like the overall visual flow direction. It breaks new ground by allowing flexible control of the generation process via three types of inputs: A) prompts, B) wireframes, and C) visual flows. The designer can provide any combination of these as input at any level of detail and will get a diverse gallery of low-fidelity solutions in response. The unique benefit is that large design spaces can be explored rapidly with very little effort in input specification.

The model allows designers to input any combination of A) textual prompts, B) wireframes, and C) visual flow direction controls. The model rapidly explores a large number of low-fidelity GUI design alternatives. Here are samples generated from different input combinations.

LEARNING TO SIMULATE USERS

While computational representations that capture essential design domain knowledge are valuable for UI design, evaluating these interfaces often requires expensive usability testing. To address this challenge, I developed models that learn to simulate interactive behaviors, enabling adaptive user interfaces that better serve users by considering specific abilities and needs. My research addresses this gap through two main areas: modeling user visual engagement with UIs to simulate perception, and developing vision-language agents for dynamic UI interaction, particularly in accessibility testing. By simulating user behaviors, I aim to **align interface performance more closely with user expectations**, creating interfaces that intuitively respond to user behaviors.

Reinforcement Learning for Visual Engagement with UIs. Understanding how users interact with and perceive UIs is essential for aligning design intent with user experience. Traditional methods, like visual attention studies, offer valuable insights for UI evaluation but can be time-intensive. In contrast, modeling eye-tracking provides a more efficient approach by revealing where users focus their attention, helping designers optimize layout and hierarchy. Thus, I collected a large-scale eye-tracking dataset across various UI categories [6] and proposed a Transformer-guided deep reinforcement learning (RL) model to predict user scanpaths [9]. The model estimates where users are likely to direct their gaze, capturing sequences of fixations that inform design improvements. By generating personalized scanpaths based on individual user data, the model offers a deeper understanding of viewing preferences. Inte-



I created models to predict saliency maps and scanpaths to simulate viewing behaviors.

grating such personalized scanpath predictions into the design process enables the creation of UI layouts adaptive to individual preferences. Personalized designs, tailored to user-specific gaze patterns, outperform generic layouts in capturing attention and enhancing user engagement.

Vision-Language Agent for UI Interaction. In the era of advanced language models, the potential to enhance UI interactions through AI has never been greater. While traditional large language models (LLMs) excel in processing text, their effectiveness is limited by a lack of understanding of the visual components inherent in UIs. By integrating recent advancements in vision-language models (VLMs) with UI interaction, I am paving the way for the development of an interactive agent in the UI domain. I proposed ILuvUI, a vision-language agent designed for UIs [11]. This agent can understand UI elements and their structure, answer questions about UIs, and simulate human interactions for specific tasks. ILuvUI excels in various UI tasks, such as UI transition predictions and goal-based planning. It not only understands and interacts with UIs but also automates UI processes. For example, it can generate and verify task steps based on the current UI screen. This automation is especially useful in accessibility testing [17], where ILuvUI verifies the effectiveness of accessibility features by comparing task execution with and without them.

RESEARCH AGENDA

Moving forward, my research will continue to develop computational models to support a wider range of users, bridging the gap between AI-driven feedback and human-centered design to advance the field of intelligent UIs.

Human-in-the-loop Generative AI for UI generation. I aim to develop models that enable flexible and diverse UI generation, allowing both experts and novices to specify requirements—such as themes, styles, sketches, color palettes, resizable layout patterns, and device specifications (e.g., mobile, laptop, XR headsets)—for guiding UI creation and adaptation. These models also predict and integrate user-specific factors, such as scanpaths, activities, abilities, and commercial considerations. Future work will focus on enabling users to input a wide range of requirements directly or enabling models to detect user states automatically, facilitating real-time UI customization that adapts to individual needs. I have already taken an initial step by developing a model that suggests cohesive color palettes for the entire UI based on user-selected modifications to a single element [16]. Moving forward, I will concentrate on incorporating more complex user inputs to develop generalized UI models that align with human behaviors, social contexts, and diverse user needs. By embedding such adaptability, these models can be used to **improve UI accessibility**, allowing, for example, designs to be adapted to users with color blindness by incorporating additional visual cues beyond color alone. This level of personalized control can be potentially achieved by capturing and simulating human behaviors or by employing human-in-the-loop methods, such as reinforcement learning from human feedback (RLHF), to ensure alignment with human expectations.

Multimodal Prediction on Human Behaviors. I will explore interdisciplinary approaches that integrate the capturing and synthesizing of human behavior and cognition. Building on my previous work—predicting gaze movement [9, 13], reconstructing 3D objects [1], capturing 3D human movements from videos [7], and using Doppler radar for activity recognition [18]—I will explore deeper into computational models that predict human behaviors, human-object interactions, and user intentions. Future work will involve developing models that not only capture but also synthesize human-like behaviors, simulating interactions that consider human goals, environment, and context. By making computers more perceptive of human intentions, this research aims to create more intuitive, adaptive, and responsive human-computer interfaces, ultimately leading to natural interactions akin to human-to-human communication. Future work will explore how multimodal inputs, such as audio, influence attention in distraction-prone environments like driving or digital interaction. This includes predicting scanpaths by integrating video, scene context, audio cues, and task relevance. Such models hold potential applications in **healthcare and education**, where gaze behavior can serve as a biosignal, providing insights into health or learning conditions, e.g., Autism diagnosis and monitoring, and supporting early detection of various cognitive and physiological states.

Adaptive Agents. With models capable of predicting the consequences of human behaviors before they occur, I will explore **interactive anticipatory agents** that respond to user intent, providing relevant information and support without explicit prompts. These agents could apply the planning and reasoning capabilities of large language models (LLMs) to deliver adaptive responses aligned with user goals, proactively addressing evolving user needs. This approach is also applicable to **mixed and virtual reality** environments, where agents could dynamically adjust to diverse 3D spaces. For example, our ORC Layout representation has been successfully applied to adapt video lectures for display on augmented reality (AR) headsets [22], optimizing the positioning of video elements in 3D. However, current methods primarily focus on spatial optimization. Future work could further design adaptive agents that proactively provide assistance based on the environment and user actions, creating responsive experiences in interactive 3D settings.

CONTRIBUTIONS TO COMMUNITY

Community Services. I have made contributions to the research community through various roles. I have served on Organizing Committees, including as Accessibility Chair for CHI 2023 and CHI 2024, and have been on different Program Committees, holding positions, such as Associate Chair for CHI 2025, Program Committee Member for IUI2023 and IUI2025, Committee Member for ASSETS 2024 Experience Reports, Workshop Juror for CHI workshop proposals, and Associate Chair for CHI Late Breaking Work for five years. Furthermore, I have participated six times in the PhD and Master's Admissions Committees.

Community Building. Community building has also been a central focus of my work. I have been significantly committed to developing the field of Computational UI. I initiated and served as the lead co-organizer for three Computational UI workshops at CHI 2022[1], CHI 2023[2], and CHI 2024[3]. These workshops brought together researchers from various HCI sub-disciplines, intersecting fields such as HCI, ML, CV, and NLP, as well as participants from different stages of their research careers, spanning both industry and academia. I encouraged discussions on the challenges and opportunities in computational approaches to understanding, generating, and adapting user interfaces. In addition, I co-organized the Mobile Eye-based Interaction Workshop at ETRA 2023[4], the largest eye-tracking-specific conference. We invited people to explore new research on visual behavior and eye-based interaction in mobile, everyday settings.

Future. Looking ahead, I will continue maintaining and expanding our Computational UI community, potentially through several initiatives. I will continue organizing workshops and events, such as workshops at CHI or related conferences such as ICML and CVPR. To facilitate deeper discussions on our research topics, I intend to coordinate one-week seminars. such as Dagstuhl seminars[5] and establish a special issue in leading journals like ACM Transactions on Computer-Human Interaction (TOCHI) to encourage more publications in this field. My vision also includes hosting a summer school on computational interaction[6] and potentially creating a summer research program that introduces undergraduate and graduate students to Computational UI and Human Behavior. This program would offer students the opportunities to collaborate on research projects under the mentorship of faculty members and research scientists working in this area.

[1] https://sites.google.com/nd.edu/computational-uichi22/home
[2] https://sites.google.com/nd.edu/computational-uichi23/home
[3] https://sites.google.com/view/computational-uichi24
[4] https://www.petmei.org/2023/
[5] https://www.dagstuhl.de/en/seminars/dagstuhl-seminars
[6] https://cixschool2022.cs.uni-saarland.de/

[1] **Yue Jiang**, Dantong Ji, Zhizhong Han, and Matthias Zwicker. "SDFDiff: Differentiable rendering of signed distance fields for 3D shape optimization." In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. **(CVPR2020 Oral Top 5%)**

[2] **Yue Jiang,** Ruofei Du, Christof Lutteroth, and Wolfgang Stuerzlinger. "ORC layout: Adaptive UI layout with OR-constraints." In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. **(CHI2019)**

[3] **Yue Jiang**, Wolfgang Stuerzlinger, Matthias Zwicker, and Christof Lutteroth. "ORCSolver: An efficient solver for adaptive UI layout with or-constraints." In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. **(CHI2020)**

[4] **Yue Jiang**, Wolfgang Stuerzlinger, and Christof Lutteroth. "ReverseORC: Reverse engineering of resizable user interface layouts with or-constraints." In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. **(CHI2021)**

[5] **Yue Jiang**, Christof Lutteroth, Rajiv Jain, Christopher Tensmeyer, Varun Manjunatha, Wolfgang Stuerzlinger, Vlad Morariu. "ORCDoc: Adaptive Documents through Optimizing both Content and Layout." *In Proceedings of IEEE Symposium on Visual Languages and Human-Centric Computing.* **(VL/HCC2024)**

[6] **Yue Jiang**, Luis A. Leiva, Hamed Rezazadegan Tavakoli, Paul RB Houssel, Julia Kylmälä, and Antti Oulasvirta. "UEyes: Understanding Visual Saliency across User Interface Types." In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. **(CHI2023)**

[7] **Yue Jiang**, Marc Habermann, Vladislav Golyanik, Christian Theobalt. "HiFECap: High-Fidelity and Expressive Capture of Human Performances from Monocular Videos." *British Machine Vision Conference.* **(BMVC2022)**

[8] **Yue Jiang**, Changkong Zhou, Vikas Garg, and Antti Oulasvirta. "Graph4GUI: Graph Neural Networks for Representing Graphical User Interfaces." In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. **(CHI2024)**

[9] **Yue Jiang***, Zixin Guo*, Hamed Rezazadegan Tavakoli, Luis A. Leiva, and Antti Oulasvirta. "EyeFormer: Predicting Personalized Scanpaths with Transformer-Guided Reinforcement Learning." *ACM Symposium on User Interface Software and Technology.* **(UIST2024)**

[10] **Yue Jiang**, Elham Sadr, Christof Lutteroth, Wolfgang Stuerzlinger. "Layout-Less-Layout: The L3-Editor, a GUI Editor for Resizable Layouts." **(in submission to IJHCS)**

[11] **Yue Jiang**, Eldon Schoop, Amanda Swearngin, Jeffrey Nichols. "ILuvUI: Instruction-tuned LangUage-Vision modeling of UIs from Machine Conversations." **(CHI2024 Workshop Paper)**

[12] Aryan Garg*, **Yue Jiang* (project advisor & co-first author)**, Antti Oulasvirta. Controllable GUI Exploration. **(In submission to CHI2025)**.

[13] Zixin Guo*, **Yue Jiang* (co-first author)**, Aini Putkonen, Luis Leiva, Antti Oulasvirta. Scanpath Prediction for Visual Search on Graphical User Interfaces. **(In submission to CVPR2025)**.

[14] Jiao Sun, Tongshuang Wu, **Yue Jiang**, Ronil Awalegaonkar, Xi Victoria Lin, Diyi Yang. Pretty Princess vs. Successful Leader: Gender Roles in Greeting Card Messages. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. **(CHI2022 Best Paper Honorable Mention)**.

[15] Yi-Hao Peng, Faria Huq, **Yue Jiang**, Jason Wu, Amanda Li, Jeffrey P. Bigham, and Amy Pavel. DreamStruct: Understanding Slides and UIs via Programmable Synthetic Data Generation. *European Conference on Computer Vision* **(ECCV2024)**.

[16] Busra Asan, Lena Hegemann, **Yue Jiang (project advisor)**, Antti Oulasvirta. "Suggesting Colors for UI Designs with Graph Neural Networks." (**CHI2024 Workshop Paper**)

[17] Maryam Taeb, Amanda Swearngin, Eldon Schoop, Ruijia Cheng, **Yue Jiang**, and Jeffrey Nichols. "Axnav: Replaying accessibility tests from natural language." In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. **(CHI2024)**

[18] Karan Ahuja, **Yue Jiang**, Mayank Goel, and Chris Harrison. "Vid2Doppler: synthesizing Doppler radar data from videos for training privacy-preserving activity recognition." In *Proceedings of the CHI Conference on Human Factors in Computing Systems*.

(CHI2021)

[19] Yao Wang, **Yue Jiang**, Zhiming Hu, Constantin Ruhdorfer, Mihai Bâce, Andreas Bulling. VisRecall++: Analysing and Predicting Visualisation Recallability from Gaze Bahaviour. *In Proceedings of the ACM on Human-Computer Interaction (**PACM-HCI**). Presented at the 2024 ACM Symposium on Eye Tracking Research and Applications* (**ETRA2024**).

[20] Parvin Emami, **Yue Jiang**, Zixin Guo, Luis A. Leiva. Impact of Design Decisions in Scanpath Modeling. In *Proceedings of the ACM on Human-Computer Interaction* (**PACM-HCI**). Presented at the *2024 ACM Symposium on Eye Tracking Research and Applications* **(ETRA2024)**.

[21] Aini Putkonen, **Yue Jiang**, Jingchun Zeng, Olli Tammilehto, Jussi P. P. Jokinen, Antti Oulasvirta. Understanding Visual Search in Graphical User Interfaces. *International Journal of Human-Computer Studies*. **(IJHCS Minor Revision)**.

[22] Han Xiao, Ashwin Ram, **Yue Jiang**, Rui Yao, Youqi Wu, Bowen Wang, Wolfgang Stuerzlinger, Shengdong Zhao. “AdaptVid: Adapting Video Lectures to Head-Mounted Displays.” (**In submission to IMWUT2025**)

[23] Parvin Emami, **Yue Jiang**, Antti Oulasvirta, Luis A. Leiva. User Interface Optimization with Hierarchical Reinforcement Learning. **(To be submitted to Siggraph2025 conference track)**