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 (HCI), computer vision (CV), and deep learning (DL), 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 develop AI models to assist UI designs that anticipate user intentions, personalize interactions, and enhance goal achievement. These computational approaches enable UIs to dynamically adapt to users and contexts, promoting effective human-AI collaboration.

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. Specifically, these representations 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, allowing for expert guidance while ensuring that users retain control over their interactions [8, 12, 16, 23].

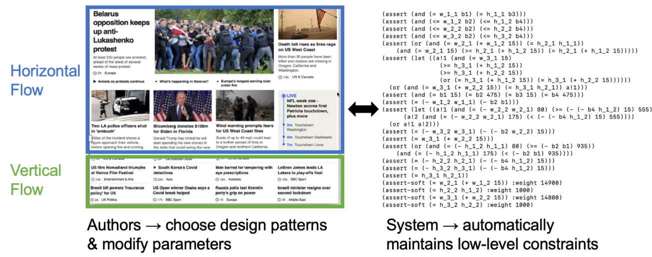
Moreover, I develop models that **simulate and predict human behaviors to facilitate automatic personalized adaptation**. These models encompass various human behaviors, including eye movements [6, 9, 13, 19, 20, 21], human motion (reconstruction from images, videos, or sensing modalities) [1, 7, 18], and interactions (simulated by interactive agents that could respond to user requirements) [11, 14, 15, 17]. Simulating these behaviors enables personalized optimization and enhances the AI to adaptively respond to user needs.

I have published my work in **top-tier Human-Computer Interaction venues such as CHI, UIST, and CSCW, leading Computer Vision 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).

COMPUTATIONAL REPRESENTATION

Advancing UI research requires moving beyond generic data-driven approaches to representations embedding domain-specific knowledge. A key challenge is to **capture 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 representations are essential for understanding, manipulating, and optimizing UIs by capturing unique explanatory factors. Structural properties—such as alignment, grouping, and layout constraints—remain difficult for current deep learning methods to capture. Effective representations capture human insight, allowing improved model development to assist humans. Computational methods can use such representations to identify and disentangle underlying explanatory factors, ultimately enhancing the computation of UIs.

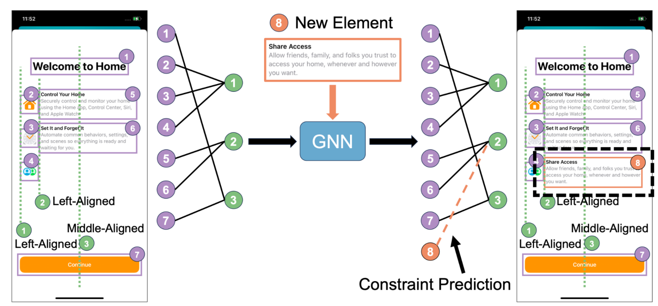
Constraint-Based Optimization for Adaptation to Contexts and User Preferences. As user needs diversify, designing adaptive UIs for different contexts is increasingly challenging. Through an elicitation study, I found that designers often manually create multiple layouts for different devices and needs, which are tedious and error-prone processes [10]. Despite advanced design tools, managing numerous parameters and layout models remains complex, making designers prefer to rely on manual adjustments—this exacerbates their workload and raises the risk of errors. To address this, in my early work, I developed the OR-Constraint (ORC) Layout representation, **automating UI adaptation to maintain design consistency without requiring intervention from a designer or programmer** [2]. ORC Layout proposes disjunctive constraints for complex adaptive design patterns, allowing UI elements to adapt across screen sizes and incorporate user preferences through automatic content generation, resulting in flexible layouts for authors and end-users [5]. This reduces the need for manual synchronization, minimizes errors, and enables designers to focus on high-level design patterns instead of low-level constraints. However, constraint-based optimization still often requires extensive manual setup for initial designs and lacks actionable suggestions for users.



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

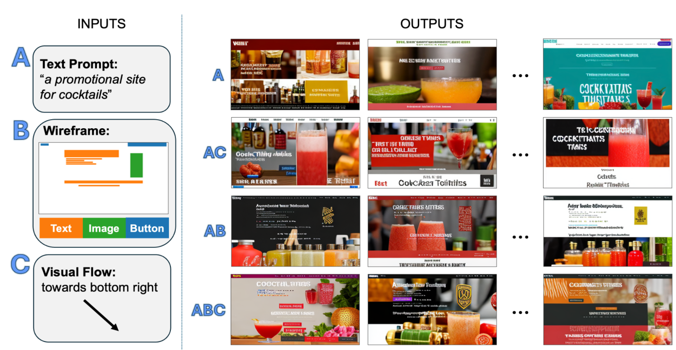
Graph Neural Networks for Human-in-the-Loop AI Assistance. Current deep learning methods can generate UI designs directly from data but limit designers' control, especially making it challenging to modify the generated UIs. My objective is to create a representation that **maximally exploits domain knowledge, using stable, universal UI design characteristics without learning from scratch, while also adapting to design-specific tendencies without manual specification**.

I explored how to represent universal layout principles as constraints in a graph, reducing the need for learning them from scratch [8]. In the graph, UI elements are represented as nodes connected to constraint nodes capturing essential design principles, like alignment and grouping, encompassing both UI element properties and layout constraints. At the same time, genre-specific tendencies are learned by using graph neural networks (GNNs) to capture the design features unique to UIs, improving how relationships between UI elements and layouts are understood. This approach bridges the gap between AI capabilities and domain knowledge, enabling an iterative and collaborative design process. It enhances the UI design workflow by preserving designers' creative control while offering AI-driven feedback and efficiency. The model supports UI autocompletion by predicting the size and position of unplaced elements with 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 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. The graph comprises element nodes (purple) 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 offer creative control in UI design but also struggle with unprecise inputs. During the early stages of UI design, designers need to create many sketches to explore ideas. Existing tools often fail here, demanding excessive detail, while generative AI has difficulty translating loose concepts from prompts. To address this, I proposed a diffusion-based approach to generate interface sketches with minimal effort [12]. My insight was that we need to have a neural representation that can **take flexible inputs and produce diverse outputs**. Thus, I integrated a diffusion model with two adapters—modular mechanisms that provide targeted guidance during image generation. One adapter controls local properties (UI element positioning and type), and the other manages global properties (e.g., overall visual flow direction). This approach allows ground-breaking flexible control over the generation process using three input types: A) prompts, B) wireframes, and C) visual flows. Designers can input any combination of these at varying levels of detail, generating a diverse gallery of low-fidelity solutions quickly and with minimal effort.

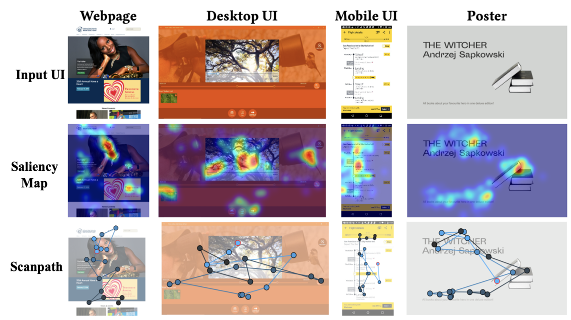


The model allows designers to input any combination of A) textual prompts, B) wireframes, and C) visual flow direction controls. The model rapidly explores a large number of low-fidelity GUI design alternatives. Here are samples generated from different input combinations.

LEARNING TO SIMULATE USER BEHAVIORS

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, I developed models that simulate user behaviors, enabling adaptive user interfaces that cater to specific user abilities and needs. My research focuses on two areas: modeling user visual engagement with UIs to **simulate perception** and developing vision-language agents to **simulate 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 to align design intent with user experience. Traditional methods, like visual attention studies, provide valuable insights for UI evaluation but can be time-consuming. In contrast, eye-tracking modeling is more efficient, 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 will likely direct their gaze, capturing fixation sequences that inform design improvements. By generating personalized scanpaths for individual users, the model enhances understanding of viewing preferences. Integrating such personalized scanpath predictions into the design process enables the creation of UI layouts adaptive to individual viewing preferences, outperforming generic layouts in capturing attention and enhancing user engagement.



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

Vision-Language Agent for UI Interaction. In the era of advanced language models, AI's potential to enhance UI interactions has never been greater. Although traditional large language models (LLMs) excel in processing text, they struggle with understanding the visual components of UIs. By integrating recent vision-language models (VLMs) with UI interaction, I developed an interactive agent for the UI domain. I proposed ILuvUI, a vision-language agent designed for UIs [11]. This agent can understand UI elements and structures, answer UI-related questions, and simulate human interactions for tasks. It excels in UI tasks like UI transition predictions and goal-based planning, automating processes such as generating and verifying task steps based on the current UI screen. This is particularly beneficial in accessibility testing [17], where ILuvUI verifies accessibility features by comparing task execution with and without them.

FUTURE RESEARCH DIRECTIONS

Moving forward, my research will continue to develop computational models to support a wider range of users, tasks, and contexts, 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 support 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 UI creation and adaptation. These models will predict and integrate user-specific factors, including scanpaths, activities, abilities, and commercial considerations. Future work will enable users to directly input a broad range of requirements or allow models to detect user states automatically, facilitating real-time UI adaptation for individual needs. I have 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 incorporate more complex user inputs to create generalized UI models that align with human behaviors, social contexts, and diverse user needs. By embedding adaptability, these models can **enhance UI accessibility** by, for example, incorporating additional visual cues for users with color blindness. Personalized control could be achieved through 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 to capture and synthesize 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 aim to deepen computational models that predict human behaviors, human-object interactions, and user intentions. Future work will focus on models that not only capture but also simulate human-like behaviors, considering human goals, environment, and context, to create more perceptive and adaptive interfaces that approach the naturalness of human-to-human interaction. I will also examine how multimodal inputs, such as audio, influence attention in distraction-prone environments like driving or digital interaction, including predicting scanpaths by integrating video, scene context, audio cues, and task relevance. Such models have potential applications in **healthcare and education**, where eye movement as a biosignal could provide insights into conditions such as Autism diagnosis or support early detection of various cognitive and physiological states.

Adaptive Agents. Building on models that predict human behaviors, I will develop **interactive anticipatory agents** that respond to user intent, delivering relevant information and support without explicit prompts. These agents could apply the planning and reasoning capabilities of LLMs to offer adaptive responses aligned with user goals, proactively addressing evolving user needs. This approach also extends to **mixed and virtual reality** environments, where agents could adjust dynamically to 3D spaces. For instance, our ORC Layout representation has been applied to adapt video lectures for augmented reality (AR) headsets [22], optimizing video element positioning in 3D. However, current methods largely emphasize spatial optimization. Future work will design adaptive agents that provide proactive assistance based on the environment and user actions, enhancing interactivity and responsiveness in 3D settings.

CONTRIBUTIONS TO COMMUNITY

Community Services. I contributed to the research community through various roles, including serving on Organizing Committees as Accessibility Chair for CHI 2023 and CHI 2024, and on Program Committees, as Associate Chair for CHI2025, 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. Additionally, I have participated in the PhD and Master's Admissions Committees six times.

Community Building. Community building has been a core focus of my work, particularly in advancing 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], gathering researchers from HCI, ML, CV, and NLP and various career stages across industry and academia. These workshops encouraged discussions on computational approaches for understanding, generating, and adapting user interfaces. Additionally, I co-organized the Mobile Eye-based Interaction Workshop at ETRA 2023[4], the largest eye-tracking conference, inviting people to explore new research on visual behavior and eye-based interaction.

Future. Looking forward, I aim to further expand the Computational UI community by organizing workshops and events at CHI or related conferences like ICML and CVPR. I plan to organize one-week seminars. such as Dagstuhl seminars[5] and initiate a special issue in leading journals like ACM Transactions on Computer-Human Interaction (TOCHI) to promote publications in this field. My vision also includes launching a summer school on computational interaction[6] and potentially a summer research program to introduce undergraduate and graduate students to Computational UI and Human Behavior. This program would offer students the opportunities to collaborate on research projects with mentorship from 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." **(CVPR2020 Oral Top 5%)**

[2] **Yue Jiang,** Ruofei Du, Christof Lutteroth, and Wolfgang Stuerzlinger. "ORC layout: Adaptive UI layout with OR-constraints." **(CHI2019)**

[3] **Yue Jiang**, Wolfgang Stuerzlinger, Matthias Zwicker, and Christof Lutteroth. "ORCSolver: An efficient solver for adaptive UI layout with or-constraints." **(CHI2020)**

[4] **Yue Jiang**, Wolfgang Stuerzlinger, and Christof Lutteroth. "ReverseORC: Reverse engineering of resizable user interface layouts with or-constraints." **(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." **(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." **(CHI2023)**

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

[8] **Yue Jiang**, Changkong Zhou, Vikas Garg, and Antti Oulasvirta. "Graph4GUI: Graph Neural Networks for Representing Graphical User Interfaces." **(CHI2024)**

[9] **Yue Jiang***, Zixin Guo*, Hamed Rezazadegan Tavakoli, Luis A. Leiva, and Antti Oulasvirta. "EyeFormer: Predicting Personalized Scanpaths with Transformer-Guided Reinforcement Learning." **(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. **(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. **(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." **(CHI2024)**

[18] Karan Ahuja, **Yue Jiang**, Mayank Goel, and Chris Harrison. "Vid2Doppler: synthesizing Doppler radar data from videos for training privacy-preserving activity recognition." **(CHI2021)**

[19] Yao Wang, **Yue Jiang**, Zhiming Hu, Constantin Ruhdorfer, Mihai Bâce, Andreas Bulling. "VisRecall++: Analysing and Predicting Visualisation Recallability from Gaze Bahaviour." (**ETRA2024**)

[20] Parvin Emami, **Yue Jiang**, Zixin Guo, Luis A. Leiva. Impact of Design Decisions in Scanpath Modeling. **(ETRA2024)**.

[21] Aini Putkonen, **Yue Jiang**, Jingchun Zeng, Olli Tammilehto, Jussi P. P. Jokinen, Antti Oulasvirta. Understanding Visual Search in Graphical User Interfaces. **(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)**