Ryan Louie — *Research Statement*

My research pushes the boundaries of human-AI interfaces to create new AI-enabled experiences that are socially connecting and personally empowering. To do this, my work develops a critical ***human-AI interface layer*** on top of pretrained AI capabilities to empower less-technical designers and end-users to effectively communicate human concepts for desired experiences and behaviors so that AI systems can better facilitate them in AI-enabled applications. My work broadly supports creators and end-users interfacing with AI capabilities, and advances a world in which AI technologies are more accountable and inclusive to people. My research sits at the intersection of the fields of HCI, social computing, and human-AI co-creation, and applies theories and methods from machine learning, decision-science, and design and creativity studies.

While AI capabilities for detection and generation have advanced significantly over the last decade, people's ability to interface with AI systems in ways that support their desired human goals has not kept up. First, although AI detectors are integral for applications that facilitate opportunities for experiences, creators spend more time programming common functions for monitoring and coordinating, instead of on structuring the experiential aspects for each application. Second, although AI detectors can be used to define applications that trigger in response to low-level events (e.g., located at a restaurant serving *soup*), it remains difficult for designers to use these detectors to express their higher-level concepts of situations that support human experiences (e.g., places where loved ones can enjoy a warm meal on a cold day together). Third, although an AI system can execute a desired experience, it can fail to encode human expectations for its execution and consider whether these expectations can be satisfied, which leads to disappointing experiences when an AI is actually deployed. Fourth, while AI capabilities for generation can power new tools for co-creation, novices lack interfaces for steering generative outputs in order to have greater control during the co-creation process.

In summary, interfacing with AI capabilities is challenging because designers and end-users cannot easily (1) structure experiences; (2) express higher-level concepts; (3) encode human expectations for execution; and (4) steer generative outputs. Through my dissertation work, and through two internship projects at Google Research, I have developed new human-AI interfacing technologies along each of these four dimensions and used them to enable new AI-powered applications in two domains: (A) a new context-aware platform for enabling opportunistic social experiences that help friends and family connect across distance; and (B) new generative music tools for empowering novices to co-create with generative AI.

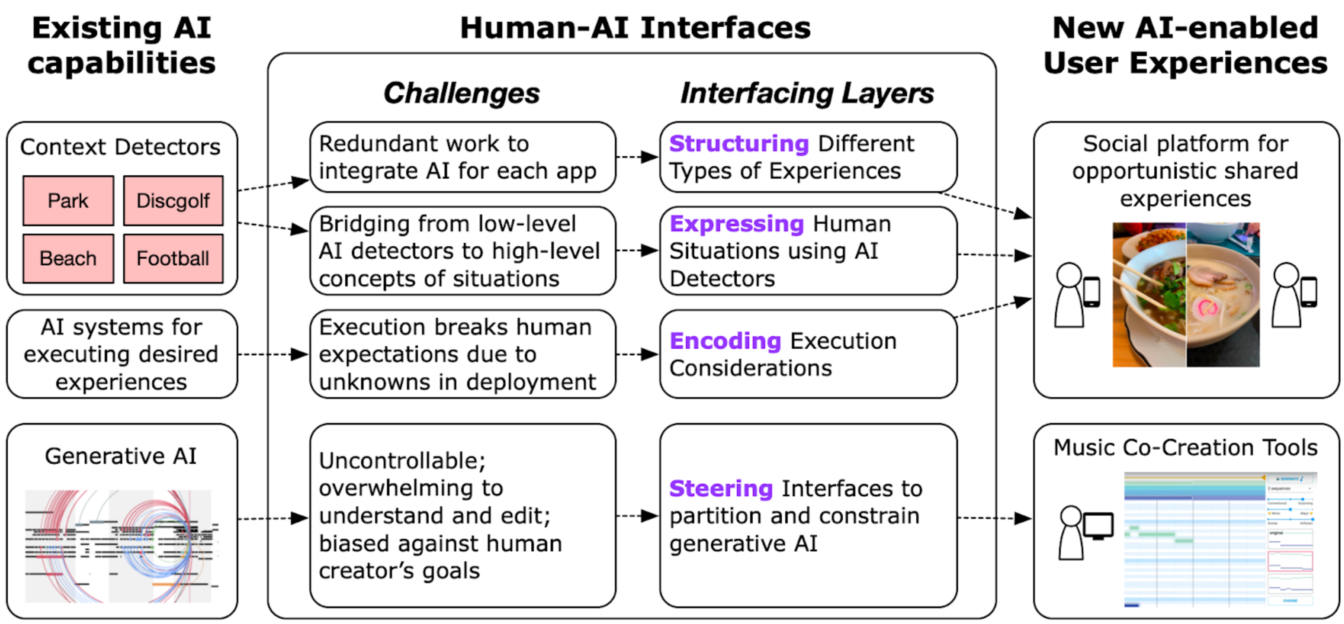

Figure 1: Human-AI Interfaces support designers and end-users to more easily interface with existing AI capabilities in order to create and realize new AI-enabled user experiences.

Human-AI Interfaces for Enabling Opportunistic Social Experiences

My dissertation work creates new human-AI interfacing layers that enable a novel context-aware social platform called Opportunistic Collective Experiences (OCEs). Unlike existing social media platforms for posting and passively engaging in feeds, OCEs identify shared situations and structure shared activities in those situations to make actively engaging easier amongst friends and family who live across distance. For example, an OCE might structure a shared experience between two friends who are in similar situations (e.g., remote friends visiting cafés in their respective cities raising their beverages for a cheers) or an interdependent activity for engaging across distributed contexts (e.g., people co-creating a storybook with friends who each contribute a unique story page based on an opportune situation they encounter). Enabled by human-AI interfaces, I built a computational platform for (1) structuring opportunities for shared experiences and activities enabled by AI systems, (2) expressing high-level concepts of situations and experiences using low-level AI detectors, and (3) encoding human expectations for execution to ensure satisfying experiences after AI handoff. **In my CSCW '20 paper [4], I conducted a 3 week deployment study with geographically-distant college alumni (N=28)** and found that OCEs that were created using the platform led to people feeling higher levels of being part of a group ($p = 0.03$) and of sharing similar experiences and engaging in joint activities ($p = 0.0001$) than posting and replying on social media feeds. Users noted that OCEs also reduced the activation energy required to initiate a contact with someone because the moments and activities it facilitates gives a common reason for interacting.

Structuring opportunities for shared experiences and activities enabled by AI systems

To enable OCEs, my first human-AI interface contribution is a structuring layer that allows creators to focus on describing the experiential aspects of their application in a few lines, while abstracting out the routine work of interfacing with detection and coordination capabilities for each AI context-aware application. To do this, the structuring interface provides an API language for implementing different types of OCEs for connecting in opportune moments across distance (see Figure 2), and an underlying opportunistic execution engine that automatically monitors users' contexts, sends opportunities when their situations match, and progresses the stages of an experience after contributions are made.

A key benefit of having a structuring layer is it easily allows for multiple applications using similar AI capabilities to be created with little effort. Already using the OCE platform, my students and I have designed and prototyped new experiences for actively engaging during opportune moments, including *Situated Self-Disclosures* for connecting with a new acquaintance [3], *Collective Narratives* that highlight a community's mutual experience despite being social distant (e.g., college senior's final week during the pandemic) [2], and a *collaborative storytelling experience* that casts users in character roles based on their situation (e.g., a geographically-distributed Murder Mystery) [1]. **The undergraduate and master students that I mentored on these projects were awarded 2nd place at the ACM CHI'22 undergraduate research competition and published Extended Abstracts at CHI'20 and CSCW'21.**

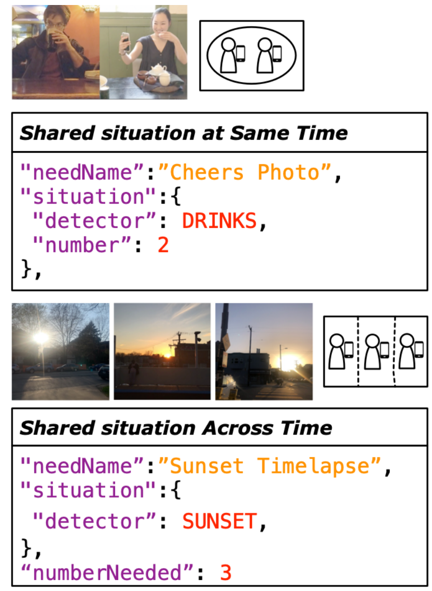

Figure 2: The structuring layer allows creators to focus on describing the experiential differences of their application in a few lines, while abstracting out the routine work of interfacing with similar detection and coordination capabilities.

Expressing high-level concepts of situations and experiences using low-level AI context detectors

Beyond structuring, my work also develops human-AI interfaces for designers to express their high-level concepts of situations, like "situations to toss a frisbee" using low-level AI context detectors, like *park*, *soccer field*, *beaches*. While advances in machine sensing and machine learning have led to a wide-availability of detectors of human context, available AI detectors are limited to recognizing low-level contexts and locations (e.g., *parks*), but do not directly represent conceptually-rich situations where human experiences or activities could take place (e.g., "situations to toss a frisbee"). Rather than wait for new AI detectors to be trained that directly recognize high-level situations, my approach supports OCE designers in translating their concepts of a human situation into a logical expression using readily available AI context detectors. **In my CHI '22 paper [5], I developed Affinder**, a visual programming environment and human-AI interface for expression. **Affinder was designed to overcome cognitive challenges, like underscoping or concept expression inaccuracies (see Figure 3),** that designers face when expressing their high-level concepts of situations using the AIs detectors.

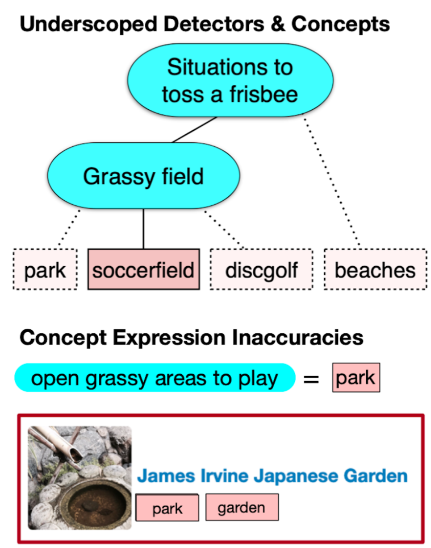

Figure 3: **Underscoping** occurs when designers fail to recall other detectors (e.g., discgolf) that apply to a concept; or when they are fixated on a type of context (e.g., Grassy field), even if other types of contexts are still appropriate for an experience (e.g., beaches). Designer's expressions may not evaluate as intended in the real-world, which creates **concept expression inaccuracies** (e.g., while one might think most parks have open grassy areas to play some like this garden really does not).

To address these cognitive challenges, Affinder provides three human-AI interfaces for expression. To resolve underscoping on detectors, Affinder provides an unlimited vocabulary search that uses textual metadata (e.g., Yelp Reviews) associated with low-level detectors which provides a rich vocabulary for querying for relevant detectors based on their conception. To address underscoping on concepts, Affinder has prompts for reflecting and expanding a designer's conception of the situation, inspired by analogical design techniques for re-representing ideas through more abstract schemas. To mitigate concept expression inaccuracies, Affinder includes simulation and repair tools for identifying misconceptions with how machine detectors operate on real-world locations. In a comparison study, I found that Affinder's human-AI interfaces for expression helped designers find relevant context-features matching their concepts, stretch their concepts for a situation beyond their initial ideas, and identify and resolve issues with the precision of their expressions on real use-cases. In addition to publishing this work at CHI'22, **I was recognized through a Google PhD Fellowship in HCI for continuing to advance human-AI interfaces for expressing human situations using AI detectors.**

Encoding human expectations for execution to ensure satisfying experiences after AI handoff

Just because an AI system knows what a desired experience should look like does not guarantee that the AI can feasibly realize the experience. Due to the opportunistic nature of people becoming available in situations to participate in OCEs, there is uncertainty over whether an AI system can successfully coordinate the completion of experiences that depend on others' availability to participate. For example, a half-half photo experience can disappoint if the AI system starts and engages the first user but never finds friends that are available to reciprocate. Without accounting for additional considerations that come with executing in the real-world with people—such as uncertainties over others' availability to participate—an AI system will not adapt its decisions to satisfy human expectations for its execution.

In a paper submitted to CHI'23 [6], I developed human-AI interfaces for encoding human expectations for execution to ensure more satisfying experiences when AI systems are deployed. The goals of these human-AI interfaces were to (1) make experiences more inclusive to users across different cities; (2) ensure that any experience the AI launched would be likely to complete; and (3) execute a partial experience that still captures the essence of the experience but with fewer people, when an experience has not yet completed. To make experiences more inclusive to users across different cities, I developed a programming environment that uses computational models of people's availability to inform designers of an experience's feasibility for different cities. In formative studies, I found this helps designers to reformulate, for example, an experience situated at "train stations" to be available to users who live in cities without train infrastructure. To ensure that experiences that are started are likely to complete, the AI system uses a decision-theoretic approach to deciding whether to activate an experience by modeling the likelihood and costs of incompletions. In a simulation study, this approach can avoid starting an experience in moments where a second person is unlikely to complete and reciprocate an experience in time. To execute satisfying partial experiences, the AI system encodes values over partial results to use the few people that are available to participate to best approximate the goals of the interaction. In deployment studies, this approach can execute a Sunset Timelapse OCE that shows the process of the sunsetting by strategically collecting photographs across the timeline (opportunistically) rather than by collecting photos from a user when they are first available. **Together, these human-AI interfaces allow AI systems to encode the human ideas for how social experiences are expected to run in order for the AI to actually facilitate these experiences effectively.**

Empowering Novices when Co-creating Music with Generative AI

During two internships at Google Research. I developed **AI steering interfaces for generative models for music,** a kind of human-AI interfacing layer that sits between a music editor application and the generative AI capabilities. I found that steering interfaces empower novices to feel *greater control and ownership* and to *effectively express their emotions* when co-creating music with a generative AI model.

In my first internship project, published at CHI'20 [7], my collaborators and I developed steering interfaces to overcome challenges novices face when interfacing directly with a deep generative infilling model. In a needfinding study, I found that novices struggled to evaluate or edit the music when ML models generate too much content at once. Others wanted to go beyond randomly "rolling dice" to generate a desired sound, and sought ways to control the generation based on relevant musical objectives. To improve the co-creation partnership, we developed COCOCO, a music editor web-interface that includes a set of AI Steering

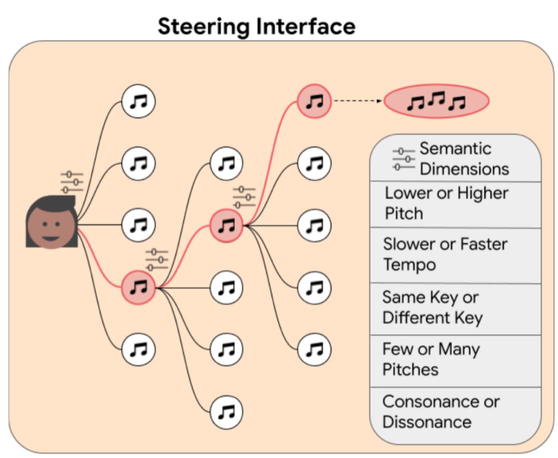

Figure 4: Steering Interfaces allow creators greater control over the co-creation process by partitioning generated outputs to be easier to understand and edit, and constraining the outputs along semantically-meaningful dimensions

Interfaces that allows users to restrict generated notes to particular voices and "nudge" outputs in semantically-meaningful directions. For example, a user could choose to generate a single accompaniment in the bass voice, and steer the generation to sound more sad. In a comparison study, I found that steering interfaces helped novices increase their control, creative ownership, and sense of collaboration with the generative ML model. Steering Interfaces allowed users to compose the song bit-by-bit, and control and understand each part more than if the AI generated it all at once. Through this work, we demonstrated that **having a human-AI interfacing layer that allows novice musicians to partition and constrain how an existing ML model generates content significantly improved composers' creative experience and their partnership with the AI.**

In my second internship project, published at IUI'22 [8], I developed AI steering interfaces for autoregressive generative ML**,** and studied how (1) steering interfaces, and (2) generative ML with greater coherence, can help novice composers' express emotions in generated music, as judged by outside listeners**.** My results shed light on how objectives of a pretrained model, such as stronger coherence, made it more difficult to generate music evoking certain emotions such as "fear"; but how the addition of steering interfaces can help to mitigate model biases by creating samples that are less likely from the model, but more aligned with the user's expression and musical goals. **This work shows how steering interfaces empower users to evoke emotions despite misalignments between their desired emotions and the generative AI's biases.**

Future Research Agenda

It's never been a more important time to be doing human-AI interfacing work. As advances in AI show increased fluency and more intelligent responses to our natural inputs, it is easy to be enamored by AI capabilities. However, doing so can cause us to create "AI-first" versions of applications that limit the ways humans are supported to connect and create. To remedy this, designers and end-users of AI applications need more effective ways to interface with existing AI capabilities to empower people in connecting with one another and in creatively expressing themselves, and to ensure AI-enabled applications execute in the real-world while respecting human expectations and safety. An explosion of progress in AI multimodal capabilities has improved the "one-way" communication of AI interpreting natural human inputs and language. To complement this progress, my future work will optimize people's capacities to interface with AI by creating modalities for human-AI dialogue. Through a back-and-forth process, such interface modalities can enable both people and AI to form mutual understanding of the other, and align the behaviors of AI systems with human expectations.

Using AI knowledge models to provide cognitive interfacing support to designers expressing concepts to AI systems. My earlier work created interfaces for expressing high-level situations an AI-system should recognize and act upon. However, just letting people express what they want is not enough. Complex expressions are hard to think about, and humans can get them wrong such as conceptions being impoverished or inaccurate. This highlights a need for cognitive interfacing support that corrects conceptual biases or misconceptions when humans are expressing or teaching concepts to AI systems. Here, I propose to use AI models that encode world knowledge (common-sense knowledge models such as COMET-ATOMIC [11]; large language models (LLMs) [12] such as GPT-3) to complement and expand a designer's own conceptions by surfacing diverse conceptions of a situation, or possible hindrances that a human designer can struggle to recall. I am excited to explore how such cognitive interfacing tools can be generally useful for expression, especially as machine teaching and few-shot prompting paradigms become more common ways to express concepts to AI systems.

New AI-mediated assistive technologies for social connection and social support. Focusing on human-AI interfacing enabled me to build new AI enabled social experiences that made it easier to connect with friends and families across distance. I am excited to uncover the needs of a variety of users who are not yet proficient at communicating, relating to, and supporting others, and explore how new AI-enabled applications could assist them with these social skills. As I create new AI-mediated assistive technologies [9, 10] for social connection and social support, taking a human-AI interfacing perspective will be crucial. In an ongoing collaboration, I aim to improve online-peer support conversations by designing an AI-system that helps novice support providers, by recognizing their ineffective strategies during real-time conversation, prompting reflections, and generating suggestions for improving the ways they understand and provide conversational support to their peers. When designing this AI-enabled application, a core interfacing challenge will be using current language AI capabilities to express empirically-informed support strategies [13] so applications can recognize ineffective strategies during conversation and generate alternatives which help-providers can learn from. Beyond interfaces needed by application designers, I will also study interfacing considerations for help providers. While AI models like LLMs might be capable of generating real-time suggestions for effective conversational responses, pilot studies could show that help-providers struggle to feel agency in the conversational process when directly using the AI's generated suggestions, which may suggest a need for new interfaces to manage the desired partnership and type of AI-mediated support.

References

[1] Gabriel Caniglia. Cast: A Context-Aware Collaborative Storytelling Platform. 2020. *In Extended Abstracts of the ACM CHI Conference on Human Factors in Computing Systems (CHI 2020).* [**mentored by me**]

[2] Nina Cong, Kevin Cheng, Haoqi Zhang, **Ryan Louie**. 2021. Collective Narrative: Scaffolding Community Storytelling through Context-Awareness. *In Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing (CSCW'21 Companion), 2021.* [**mentored by me**]

[3] Cindy Hu. Self-Disclosure for Early Relationship Development through Situated Prompts in Opportunistic Collective Experiences. 2022. *In Extended Abstracts of the ACM CHI Conference on Human Factors in Computing Systems (CHI 2022). 2nd Place in Undergraduate Student Research Competition.* [**mentored by me**]

[4] **Ryan Louie**, Kapil Garg, Jennie Werner, Allison Sun, Darren Gergle, Haoqi Zhang. 2020. Opportunistic Collective Experiences: Identifying Shared Situations and Structuring Shared Activities at Distance. *In Proceedings of ACM Human-Computer Interaction (CSCW 2020).*

[5] **Ryan Louie**, Darren Gergle, Haoqi Zhang. 2022. Affinder: Expressing Concepts of Situations that Afford Activities using Context-Detectors. *In Proceedings of ACM CHI Conference on Human Factors in Computing Systems (CHI 2022).*

[6] **Ryan Louie**, Kapil Garg, Darren Gergle, Haoqi Zhang. 2023. Flexible Strategies for Structuring and Coordinating Opportunistic Social Interactions. *Under Review at ACM CHI Conference on Human Factors in Computing Systems (CHI 2023)*

[7] **Ryan Louie**, Andy Coenen, Cheng-Zhi Anna Huang, Michael Terry, Carrie Cai. 2020. Novice-AI Music Co-Creation via AI-Steering Tools for Deep Generative Models. *In Proceedings of ACM CHI Conference on Human Factors in Computing Systems (CHI 2020), 2020.*

[8] **Ryan Louie**, Jesse Engel, and Cheng-Zhi Anna Huang. 2022. Expressive Communication: Evaluating Developments in Generative Models and Steering Interfaces for Music Creation. *In the 27th International Conference on Intelligent User Interfaces (IUI '22).*

[9] Paul Ruvolo, **Ryan Louie**, and Eric Jerman. 2023. The Cane Game: An Educational Tool for Orientation and Mobility. *Under Review at Case Studies of the ACM CHI Conference on Human Factors in Computing Systems (CHI 2023)*

[10] Chris Yoon, **Ryan Louie**, Jeremy Ryan, MinhKhang Vu, Hyegi Bang, William Derksen, Paul Ruvolo. Leveraging Augmented Reality to Create Apps for People with Visual Disabilities: A Case Study in Indoor Navigation. *In Proceedings of ASSETS 2023* in Pittsburgh, PA 2019
Best Artifact Award

[11] Jena D. Hwang, Chandra Bhagavatula, Ronan Le Bras, Jeff Da, Keisuke Sakaguchi, Antoine Bosselut, and Yejin Choi. "(COMET-)ATOMIC 2020: On symbolic and neural commonsense knowledge graphs." In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, no. 7, pp. 6384-6392. 2021.

[12] Fabio Petroni, Tim Rocktäschel, Sebastian Riedel, Patrick Lewis, Anton Bakhtin, Yuxiang Wu, and Alexander Miller. 2019. Language Models as Knowledge Bases?. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pages 2463–2473, Hong Kong, China. Association for Computational Linguistics.

[13] Tianying Chen, Kristy Zhang, Robert Kraut, Laura Dabbish. 2021. Scaffolding the Online Peer-support Experience: Novice Supporters' Strategies and Challenges. *Proceedings of the ACM Human-Computer Interaction (CSCW 2021)*.