CS PhD openings @ Harvard HCI's Variation Lab advised by Prof. Elena Glassman

A little history

Founded in the fall of 2018, our small group has been a home for some truly terrific interdisciplinary HCI researchers: our postdoctoral scholars Tianyi Zhang. Ian Arawjo, and writer-scientist Katy Gero are all now CS professors, and we just graduated our first PhD student! *Dr.* Priyan Vaithilingam just became a Research Scientist in the AI/ML group @ Apple, and Siebel Scholar Chelse Swoopes, who is currently focusing on legal tech, plans to graduate soon. It's time to train new students. (And we just got a new NSF grant with Ian to fund it.)

Current primary focus

AI is powerful, but it can be wrong, contextually inappropriate, and subjectively bad. We are refining our understanding of AI-resilient interfaces [Gu et al. CHI'24, Glassman et al., arXiv] that help people use AI while (a) maintaining their ability to *notice* AI choices and (b) giving them the context necessary to evaluate whether those choices are wrong, or at least not right for them. This is critical during context- and preference-dominated open-ended tasks, like ideating, searching, sensemaking, and reading or writing text and code at scale. AI-resilient interfaces improve AI safety, usability, and utility by working with, not against, human perception, attention, and cognition. To achieve this, we derive design implications from cognitive science, even when they fly in the face of common usability guidelines, e.g., [Gu et al. UIST'25]. We also aim to support users without introducing AI features that interfere with the sometimes challenging aspects of experiences that facilitate meaning beyond just 'getting stuff done' inspired by Zhang CHI'24.

On-going and future directions

Since the public papers on our website are a lagging indicator of where our thoughts are going, here are some of our current and future interest areas that, with the right student, we could develop further:

Designing for value that's hard to measure

As technology developers, we generally need to measure the impact of innovations on outcomes. We might measure whether AI-assisted outcomes are more or less fair, or determine under what circumstances users are making less accurate decisions, or are misled. However, our technologies have impacts on *non-outcome value* as well. Ethicist Talbot Brewer refers to activities with significant non-outcome value as 'dialectical'. AI assistance is unambiguously helpful in outcome-focused activities, but it can interfere with dialectical activities by shortcutting

the time, personal effort, and/or experience that would otherwise generate greater meaning. If we cannot reliably recognize activities that have dialectical qualities, we risk making AI-powered tools that users find useless or even harmful to the value they would otherwise experience without assistance.

- Inspirational reading: Haoqi Zhang's Searching for the Non-Consequential: Dialectical Activities in HCI and the Limits of Computers (CHI'24)

AI-Resilient Interface Design

AI is powerful, but it can make both objective errors and contextually inappropriate choices. We need AI-resilient interfaces [Gu et al. CHI'24, Glassman et al., arXiv] that help people be resilient to the AI choices that are not right, or not right for them. Existing human-AI interaction guidelines recommend that interfaces include user-facing features for efficient dismissal, modification, or otherwise efficient recovery from AI choices that the user does not like. However, users cannot decide to dismiss or modify AI choices that they have not noticed, and, without sufficient context, users may not realize that some of the noticed AI choices are wrong or inappropriate. We are developing guidelines that help designers and end-user understand and weigh the challenges and benefits of designing AI-resilient interfaces.

Theory-connected systems

We have made several counterintuitively effective interfaces that raised the ceiling on what human users could comprehend and reason about by exploiting theories of human concept learning while violating common usability guidelines, e.g., Examplore, ParaLib, Positional Diction Clustering, and AbstractExplorer are all "too much" at first glance, but enable users to develop more accurate, detailed mental models of concepts than prior approaches that showed fewer concrete examples and didn't pre-compute and render many-to-many cross-example relationships.

Now, we're zooming out and looking at how we can do more theory-connected system development in general. We hope to build on and complement prior and on-going work like Mackay et al.'s Generative Theories of Interaction, *design claims* described by Resnick & Kraut's book Building Successful Online Communities: Evidence-Based Social Design (MIT Press) and design briefs for psychoactive interventions [Slovak and Munson CHI'24]. We are particularly (but not exclusively) interested in building systems that leverage theories about human attention and emotional regulation [Slovak et al. TOCHI'23], which can benefit everyone, regardless of their place on the spectrum of neurodiversity.

We also hope to develop intellectual and tool-based infrastructure to support ourselves and others do more theory-connected work across HCI.

Interaction Techniques for Long-Term Human-AI Collaboration

co-PI: Ian Arawjo (UMontreal)

As artificial intelligence (AI) systems become more powerful and self-determining, they are shifting from simple tools to collaborative partners that can work with humans on complex, long-term projects. However, current AI systems may not effectively maintain communication with humans over extended periods. They often fail to track preferences, follow instructions, and/or adapt to project-specific context. The proposed work will investigate how to design AI systems that can establish and maintain a shared understanding with humans when collaborating over long-term creative projects, and enhance human transparency and control over this understanding. Potential application areas for case studies include long-form document writing, game design, and software development. An important goal for this project is to *produce open-source tools and interfaces* that will benefit industry developers, researchers, and users of AI systems across the United States and Canada; thus, we are looking for strong candidates who can lead these efforts.

Preliminary work: Semantic Commit: Helping Users Update Intent Specifications for AI Memory at Scale (UIST'25), Imagining a Future of Designing with AI: Dynamic Grounding, Constructive Negotiation, and Sustainable Motivation (DIS'24), Antagonistic AI (arXiv'24)

Open source modules and system development

Our recent experience developing open source systems, i.e., ChainForge.ai [Arawjo et al. CHI'24], has been surprisingly successful in terms of creating research outcomes, research infrastructure, and industry impact. However, we believe that most systems will not be reused in their entirety; instead, there are novel components (modules) that can be released for ourselves and others to reuse in new combinations and applications. We hope to (1) continue working to release self-contained modules with novel functionality extracted from recently published systems and (2) use these modules to develop and deploy more robust "v2" versions of systems for longitudinal study beyond short lab-based studies, as described in the UIST'25 workshop on longitudinal interaction studies of AI systems.

Applications to law & connections with policy

possibly co-advised by Prof. Jonathan Zittrain, Berkman Klein Center For Internet & Society, Harvard Law School

Like academic writing, the practice of law can have high stakes and requires careful consideration of language (e.g., words may have very specific legal meanings). This is a place where *AI-resilient interfaces* are likely to be critical, and yet we already see interfaces being used

that may be setting law professionals up to fail. For example, Chelse Swoopes worked with a Harvard Law professor to translate these ideas from HCI into the law domain: Law is vulnerable to AI influence; interface design can help (SSRN'25). Previously, our then-postdoc Katy Gero teamed up with the Harvard Law School's Library Innovation Lab to publish the CHI'24 Best Paper titled Creative Writers' Attitudes on Writing as Training Data for Large Language Models. Now, Harvard Law Professor Jonathan Zittrain is interested in co-advising an incoming member of our lab if the fit is right. This could include one day a week at the Berkman Klein Center for a multi-disciplinary rather than interdisciplinary outcome, i.e., a CS PhD with expertise in law or policy and extra skills in communicating beyond one's own field. With these and new collaborators, we hope to design the next generation of interfaces that support the practice of law, while also considering interface design jointly with plausible effective policy design.

News Production and Consumption

There's been a longstanding interest in applying our techniques for semistructured text sensemaking to news stories. We have had preliminary work accepted at the Computation + Journalism workshop several times since 2020. We recently visited the Globe for informal needfinding, and are building connections with small innovative companies in the news industry. If you have journalism or other relevant experience, let us know!

Wider CS/AI community

The Variation Lab is one of two labs within Harvard HCI, the HCI community co-led by Prof. Gajos and Prof. Glassman. Harvard HCI is also a founding member of CHARM, our human-centered AI community, which also includes Prof. Martin Wattenberg & Prof. Fernanda Viegas' Insight + Interaction Lab and Finale Doshi-Velez's Data to Actionable Knowledge group.

If you're interested in joining us

1. Fill out this Google Form
2. Apply to Harvard SEAS and put down my name where the application asks for faculty of interest

(You can also email me and prepend [phd applicant] to the subject line, but I still might miss it!)