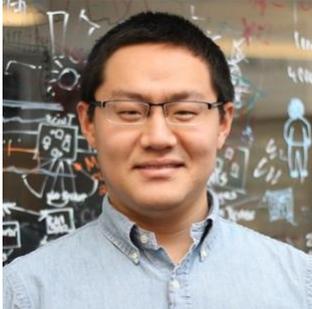


# PHD THESIS PROPOSAL

Monday, May 20<sup>th</sup>, 2019, GHC 4405, 1pm



## Anhong Guo

Human-Computer Interaction Institute

### Committee:

Jeffrey Bigham, HCII and LTI, CMU (Chair)

Chris Harrison, HCII, CMU

Jodi Forlizzi, HCII & Design, CMU

Meredith Ringel Morris, Microsoft Research Ability Group

## Human-AI Systems for Visual Information Access

The world is full of visual information that is not easily accessible. For blind people, frustrating accessibility problems because of vision are commonplace and pervasive. For example, they frequently encounter inaccessible physical interfaces in their everyday lives that are often impossible to use independently. For space owners, actionalizing camera streams into sensor data can help them better monitor, manage, and optimize the environment. Despite the advantages, these visual information are often left uncaptured.

Two trends are converging that make solving these problems tractable: AI and human computation. With recent and impressive advances, AI shows promise in understanding the visual world with computer vision. However, AI systems struggle in many real-world, uncontrolled situations, and do not easily generalize across diverse human environments. Humans, on the other hand, can be more robust and flexible in solving real-world problems that cannot be handled by AI. However, using human intelligence is slow and expensive, thus not scalable.

In my work, I investigate hybrid human- and AI-powered methods to provide robust and interactive access to visual information in the real world. They tradeoff between the advantages of humans and AI to create systems that are nearly as robust and flexible as human, and nearly as quick and low-cost as automated AI, foreshadowing a future of increasingly powerful interactive applications that would be currently impossible with either alone. In my work thus far, I have developed human-AI systems to make physical interfaces accessible for blind people. I have developed (i) VizLens, a screen reader to help blind people access static physical interfaces; (ii) Facade, a crowdsourced fabrication pipeline to automatically generate tactile overlays to appliances; and (iii) StateLens, a reverse engineering solution that makes existing dynamic touchscreens accessible. Furthermore for environmental sensing, I have developed and deployed (iv) Zensors++, a camera sensing system that collects human labels to bootstrap automatic processes to answer real-world visual questions, allowing end users to actionalize AI in their everyday lives.

To complete this dissertation, I plan to (i) harden and deploy the VizLens system to better understand how blind users of diverse vision capabilities use physical interfaces in the wild, (ii) from the deployment collect a dataset of interfaces and interface interaction, which we will then release to other researchers, and (iii) mine YouTube for videos of appliance use in order to validate StateLens at scale and release a large dataset of these real world interfaces.

Document at: [https://guoanhong.com/papers/Thesis\\_Proposal\\_AnhongGuo.pdf](https://guoanhong.com/papers/Thesis_Proposal_AnhongGuo.pdf)

