Social Sensemaking with AI: Designing an Open-ended AI experience with a Blind Child

CECILY MORRISON[*]

Microsoft Research

EDWARD CUTRELL

Microsoft Research

MARTIN GRAYSON

Microsoft Research

ANJA THIEME

Microsoft Research

ALEX S TAYLOR

City, University of London

GEERT ROUMEN

Microsoft Research

CAMILLA LONGDEN

Microsoft Research

SEBASTIAN TSCHIATSCHEK

Microsoft Research

RITA FAIA MARQUES

Microsoft Research

ABIGAIL SELLEN

Microsoft Research

AI technologies are often used to aid people in performing discrete tasks with well-defined goals (e.g., recognising faces in images). Emerging technologies that provide continuous, real-time information enable more *open-ended* AI experiences. In partnership with a blind child, we explore the challenges and opportunities of designing human-AI interaction for a system intended to support social sensemaking. Adopting a research-through-design perspective, we reflect upon working with the uncertain capabilities of AI systems in the design of this experience. We contribute: (i) a concrete example of an open-ended AI system that enabled a blind child to extend his own capabilities; (ii) an illustration of the delta between imagined and actual use, highlighting how capabilities derive from the human-AI interaction and not the AI system alone; and (iii) a discussion of design choices to craft an ongoing human-AI interaction that addresses the challenge of uncertain outputs of AI systems.

CCS CONCEPTS • Human-centered computing~Human computer interaction (HCI) • Accessibility • Empirical studies in HCI

Additional Keywords and Phrases: Human-centred AI, human-AI interaction, human-AI collaboration, intelligent systems, accessibility, blindness, visual impairment, children, disability, design

1 INTRODUCTION

Rapidly developing Artificial Intelligence (AI) technologies provide an opportunity for new types of human-AI interactions to emerge that enable people to augment, or extend, their own capabilities. People with disabilities, as early adopters of AI technologies [8], are fruitfully pushing the boundaries of how AI systems can be used and what the interaction paradigm might look like. Indeed, there are already a substantial number of AI applications that support people who are blind or low vision to achieve discrete pragmatic tasks, in which the end-to-end experience is well defined. For example, a blind user can take a photo of a short piece of text such as a conference room name, and the AI application reads out the words [47]. As AI technologies progress, there is potential to move beyond the current focus of solving discrete tasks and explore ways in which we can harness the promise of AI to assist users in richer, longer-term engagements.

In this paper, we explore a different kind of human-AI interaction through the design of an *open-ended* AI experience intended to provide children who are blind or low vision with ongoing, real-time information about other people in their vicinity (See Fig 1). An enriched understanding of social context [33,57], such as who is in the room and how they relate to the user, can empower people who are blind or low vision to proactively choose their interactions: to socialise, ask for help, or simply moderate their behavior appropriately. Yet, such social understanding is a situated, dynamic sensemaking activity driven by the user as they need it. Moreover, one's social context can change dramatically moment-to-moment as people move about. This differs from the completion of a discrete technical task, such as face recognition, as the end-to-end use scenario cannot easily be defined. This led us to consider *how we might design human-AI interaction in a system intended to support the open-ended activity of social sensemaking.*

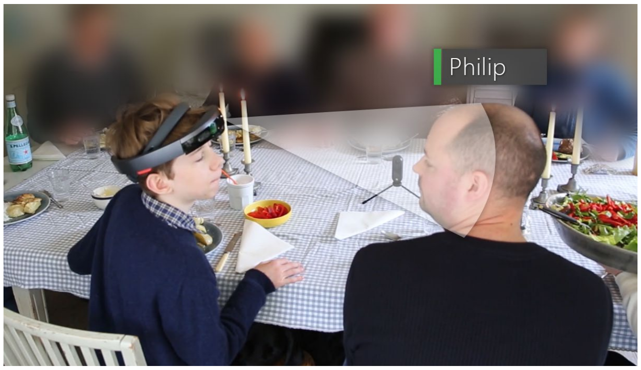

Figure 1. TH uses PeopleLens to engage with a conversation partner.

Our design exploration to support open-ended social sensemaking activities in the context of the evolving functionality of AI systems builds on the idea of *capability uncertainty* proposed by Yang et al. [64]. Following their definition, *capability* broadly refers to the functionality of the AI (ie., face recognition) and how well the system performs, including the types of errors it creates. Inherent to this however are *uncertainties* around how well a system might perform (and thus what it can do) given its probabilistic nature and dependency on context of use. Standard tools such as sketching and prototyping that help designers understand the boundaries of the technology in order to ideate, assess, and define a preferred future, or design goal, can be difficult to apply to AI systems as the capabilities often only become apparent after deployment [11,62,63].

To address this challenge, Yang et al. [64] defined a new design process which distinguishes between prototyping the AI system and prototyping the user experience of the AI system. It suggests the latter should happen through the thoughtful and choreographed deployment of an AI prototype to explore potential design decisions before the user experience is ultimately defined and evaluated. This additional part of the design process has been proposed to enable interaction designers to obtain key insights into the workings of the AI system in situ in order to respond to capability uncertainties with appropriate interaction design decisions. While a necessary step for addressing capability uncertainty in AI systems, it is more resource and time intensive than traditional rapid prototyping approaches. Consequently, our ability to learn from exemplars through research-through-design approaches may be beneficial.

Framed using a research-through-design (RtD) perspective [67], this paper describes and reflects upon the user experience prototyping phase of an open-ended AI system intended to support the social agency of blind children at school. Social inclusion in education is a known challenge for blind and low vision children, identified in the HCI literature [29], the clinical literature [41], as well as an issue observed by our research team through previous deep engagements with this community. Our goal was to provide a facility that could augment children's existing sensemaking skills, as well as scaffold proactive social interactions that allow children to experience a sense of social agency. By definition, such a system must support ongoing interaction with a child in a social milieu. Given both the technical and ethical challenges of deploying a multi-algorithm, state-of-the-art AI system that includes facial recognition in a school, we felt the responsible approach [46] was to first focus on a deep engagement with a single young blind person, his family, and his school community.

Over a period of seven months, we worked with TH, his family, and his school to respectfully craft and refine experience ideas for a system we dubbed the *PeopleLens*. TH was twelve years old at the time, and attended an independent, mainstream school. Despite strong academic abilities and the skill to travel independently, he struggled to find and interact with his friends at school. His family also described challenges in helping TH understand the power of physically orienting to social interaction, such as looking at someone when they speak. Building on previous research engagements with TH, we started design sessions in our lab, exploring potential uses of the technology. We progressively moved into more realistic settings in his school environment to understand how he used the system with his friends and teachers, and how they responded.

The user-experience prototyping discussed in this paper uses a sophisticated AI prototype [27] intended to provide people who are blind or low vision with a better understanding of their immediate social environment. The prototype uses a head-mounted augmented reality device in the form of a modified HoloLens [30] in combination with five state-of-the-art computer vision algorithms to *continuously* identify and track people in space as well as capture the gaze direction and activities of people in the vicinity. It then presents this information about the social environment to the user through spatial audio. For example, whenever the user passes their gaze over another person, they would hear the name of the person, or, if not identified, a spatialised sound to indicate the presence of a person. As this and further examples will illustrate, many design decisions are at work to translate the raw perception information provided by the AI prototype into a useful and coherent user experience.

We describe the user-experience prototyping process, the PeopleLens, and vignettes of its usage as a means to reflect upon designing with capability uncertainty in the creation of open-ended AI experiences. We show how TH, his social surroundings, and interactions with and understanding of the capabilities of the AI prototype shaped his experience and that of those around him. This situated, user-encountered capability of AI [64] is difficult to anticipate from the outset and yet fundamental to user experience design. We also discuss design decisions that address the give-and-take between user, AI system, and social setting in this ongoing, open-ended experience, speaking to the challenge of crafting coherent experiences from systems that produce uncertain and complex outputs. We make the following three key contributions:

- We describe the design and usage of the PeopleLens, an open-ended AI experience that augments social sensemaking, by a 12-year-old blind boy, offering a concrete example of how AI might enable and extend human capabilities in ways different from the discrete tasks currently supported.
- We outline and discuss the design choices that we made for the PeopleLens in order to address the capability uncertainties that arose in crafting the experience of this open-ended AI system. These highlight three issues: (i) the importance of low-density information; (ii) creating intelligibility through support information; and (iii) designing for reciprocal interaction.
- We reflect upon the PeopleLens design journey and the substantial differences between imagined and actual use, suggesting that open-ended AI systems give more scope for users to extend their own capabilities which introduces capability uncertainty derived from the human-AI interaction rather than the AI system alone.

2 RELATED WORK

We begin by situating our work and approach in the literature on human-AI interaction. We then consider current AI and Augmented Reality (AR) technologies for people who are blind and low vision and the types of interactions they afford. Finally, we consider the focus of the experience we create, supporting the social agency of children who are blind. As this literature is highly clinical in nature, we also bring in perspectives from the disability literature for balance.

2.1 Understanding Human-AI Interaction

There is a diverse literature on conceptualising and designing human-AI interaction. The HCI literature has, for the most part, rejected the conceptualisation of AI as a simulation of human capabilities despite this popular framing in machine learning communities (e.g., [49]). Emphasis has instead been put on conceptualising the (collaborative) relationship between user(s) and AI system (e.g., [21,50]). One conceptual framing that is particularly nuanced in the ways that AI systems might augment human capability is cognitive extenders [22]. It contrasts the collective intelligence of a cognitive service that an AI system might provide to a user (User + AI) with the role an AI system can play as a tool of cognitive extension (User[AI]). A cognitive extender is defined as “an external physical or virtual element that is coupled to enable, aid, enhance, or improve cognition, such that all – or more than – its positive effect is lost when the element is not present” [22]. It is this relationship of cognitive extension that we aspire to create when supporting ongoing social sensemaking with the PeopleLens.

Other researchers have focused on the *practice* of designing human-AI interaction. Guidelines have been developed to support interaction design decisions with AI systems (e.g.[4]). Experimental research has tried to consider the optimal collaboration conditions through the design and building of toy systems (e.g., [36]). Work has been done to understand the types of information users need before they begin to use an AI system, such as for medical decision-making (e.g., [10]). Not least, researchers have focused on the system side to develop new techniques to make systems more interpretable, explainable, and accountable to people (e.g., [1]). Despite the body of knowledge gained from considering human-AI interaction from multiple perspectives, a recent review paper suggests that interaction designers continue to have significant challenges in practically designing experiences with AI systems [64].

Yang et al. [64] identify two types of challenges designers face when designing AI systems: 1) uncertainty surrounding AI system capabilities in the early, or divergent stages of design; and 2) the complexity of the outputs of an evolving system in the later convergent stages. They note that in the divergent stages, for example, designers often have difficulty generating novel, purposeful uses of AI and, having done so, may have difficulty assessing the feasibility of an idea because it may depend on data or relationships to other algorithms. In the convergent stage, challenges are noted around crafting thoughtful experiences with a system that may have unpredictable behaviors. These difficulties are collectively referred to as *capability uncertainty*. They arise from the gulf between what the AI system and associated data appear to promise and the reality of what they can concretely achieve.

This paper reflects upon capability uncertainty in the concrete example of designing and using the PeopleLens. Note that while the PeopleLens does not adapt or learn over time, the complexity of the models leads to outcomes that are often difficult to predict. Variables such as lighting conditions, the density of people in space, the positioning of people and their faces in relation to a moving user, images in reflective surfaces, variability in training images for face recognition combine to create significant uncertainties in performance that are related to the concept of capability uncertainty as described by Yang, et al.

2.2 AI and AR technologies for Blind and Low Vision People

Blind and low vision people have been early adopters of AI technologies [8] and particularly computer vision-based applications. Digital systems now provide automatic image captions for images on social media platforms [61] and in operating systems, and they support exploration of visual websites, e.g., shopping for clothes [52]. Moving into the physical world, researchers have mined zebra-crossing locations to support route planning [2]; created finger-mounted cameras to read text [53] or recognise patterns on clothes [54]; and built smart phone apps to help recognise personal objects [23] and people [66]. Systems that provide access to visual information in non-digital settings have been validated with large numbers of people through crowd-supported systems (e.g., VizWiz [7]).These explorations are now being actualised in products such as SeeingAI [47], an app that uses AI to support everyday tasks such as reading signs or recognising the contents of a photo.

There is also growing research into developing audio AR experiences for people who are blind or low vision. The majority of this work focuses on navigation, providing technical contributions [9,15] or supporting users to keep a straight path using a mobile phone [39], head-mounted display [16], or an environment tracked with motion-capture technology [26]. Other research investigates the use of a wearable device to help enhance the spatial cognition of blind children [13] and support them to play and move independently in kindergarten [18]. In both cases, sound is being used to guide attention and to connect spatial understanding with proprioception to elements of the environment. So far, while these studies show positive results, these lab-based studies do not utilise any AI functionality and provide few insights into the design decisions made for, or the analysis of, the audio AR experience.

Much less research has addressed technology design to support social understanding for those who are blind or low vision. Amongst the few existing examples, social assistants have been explored that: provide the location and distance of an interaction partner and their facial expression through a tactile belt [38]; verbal presentation of emotional valence and head direction [35]; and gaze direction through tactile feedback [43]. These systems have not been deployed with users in the wild, giving us little feedback on their feasibility and the experiences offered. Yet, there is research to suggest the blind community is particularly interested in the opportunity to receive social information [33,57]. The work in this paper addresses this need.

2.3 Blind Children and Social Interaction

Responding to the specific design focus we present in this paper, we consider the literature on the challenges of social interaction for blind children and current strategies to support development. Children born blind often have substantial and ongoing difficulties with social interaction [41]. Research has shown that at school age (6–12 years), blind and low vision children show significantly poorer usage of language for social purposes compared to their sighted peers, despite good intellectual abilities and advanced linguistic skills [56]. Other difficulties include perspective taking, emotional recognition, and atypical social interaction, which can impact educational outcomes as well [41]. The extent of the difficulties is captured in the statistic that two-thirds of blind children show behavioral similarities to sighted children with autism [41].

While there is much disagreement about the underlying nature of the social challenges of blind children, there is consensus that appropriate intervention can shift the ability of blind children to relate and engage with others [20,41]. One intervention, the augmentation of toys with sound, illustrated that while it facilitated engagement, it hindered cooperative play [58]. Social skills training was also found to be effective and generalised in peer-mediated groups [45]. More typically, interaction between blind children and their sighted peers is often mediated by adults. Teachers, assistants, or parents often describe the social environment, such as who is in the classroom when a child enters, to provide the first layer of information the child needs to participate. Yet, while mediation is practical, it is problematic for social interaction as it spawns a ‘teaching assistance bubble’ [29]. This suggests an opportunity to consider how a dynamic, AI-enabled experience might support a blind child to engage with their peers.

2.4 Disability Studies Literature

Contemporary thought in the disability studies literature offers a distinctive perspective on the ways technology can work in concert with human agency. This work shows that technology is not a ‘solution’ for people with a disability by attempting to replace a physical or sensory ability a person does not have (e.g., vision or movement). Rather, assistance [25] and by extension assistive technologies [34] are appropriated to extend people’s capacities in highly personal and situationally dependent ways. This is achieved through a continuous mutual shaping of technology and person that can reconfigure the relationships between person, technology, and social context to enable new capabilities. For example, cochlear implants depend on the user’s capacity to adapt to a different way of ‘hearing’ through long-term auditory training programs, evocatively referred to by Mauldin as a “choreography of listening” [28].

It is this idea of choreography that provides a particular design stance to audio AR. It suggests a shift in emphasis from discrete outputs and efficient task completion towards approaching technology as an element in an active back-and-forth engagement with a user. There is an acknowledgement of the interdependence [31] between the many entities and activities occurring in any context and their active assembly or sensemaking by the user. This, in turn, opens up the conditions of possibility to make certain capabilities attainable. Closer to HCI, notions of interdependence [5] and an assemblage of technology to support sensemaking [57] have been put forward as framings for the development of technology that consider the often unrepresented ways people with disabilities are more than the recipients of support but, in fact, play active roles in what it means to be capable. This has been a key framing for our design work discussed in the next sections.

3 PEOPLELENS PROTOTYPE

This paper takes a research-through-design (RtD) perspective [67] to describe and reflect upon the user experience prototyping phase of the PeopleLens. In Section 3, we first document the AI prototype that we used in our design engagements, then we describe the final experience design. This provides the context for Section 4, which captures key learnings in the design journey.

3.1 AI Prototype

3.1.1 Social Sensemaking

Social sensemaking was a key driver in the creation of the AI prototype. The notion of sensemaking was originally introduced in organizational psychology [59] to describe how people come to understand the unique configuration and situated dynamics of their social surroundings that gives meaning to their collective experience over time. Weick [59] described seven properties of sensemaking that combine to help people interpret events and the social experience around them. A central component is the process of understanding who is around an individual, their physical proxemics, and the complex social information that one gleans from the communication and social history of an individual and between actors. In HCI, Russell, et al. [44] used the term *sensemaking* to capture how people develop representations and organize information to support activities such as decision-making and problem-solving. In our work, we lean on both of these concepts, using the term *social sensemaking* to describe how a person comes to understand and successfully interact with the social milieu in which they are situated.

The focus on social sensemaking derived from design ideation workshops with blind and low vision people in the UK and India [33] as well as ethnographic fieldwork with Paralympic athletes and spectators [57]. This research foregrounded both their social inclination and their skillful sensemaking strategies to understand their environment. Examples focused on where other people are in the physical space as well as who those individuals are and what activities they are engaged in. For example, a participant in India speaks of a visit to a relative's house: "Who is present today? Is cousin there?" Social sensemaking can also capture the tone of the atmosphere (e.g., tense or relaxed) and any interpersonal or group dynamics that might be at play. For example, a participant in the UK said: "A way to know how someone is responding when I'm breaking bad news [as a doctor] in a hospital context."

Sensemaking [40] more generally highlights the ways that people bring together multiple strands of information from different senses over time to understand a situation. It was observed in our ethnographic study that participants felt most confident to engage in social interactions when they were able to “work out” what was happening around them from triangulating multiple cues. For example, a blind person might expect someone to be attending a meeting and then hear that person’s unique phone ring, encouraging the blind person to approach or introduce themselves. The AI prototype was imagined as an additional layer of information to enable users to extend their existing sensemaking strategies; it is not intended to replicate the social sensemaking process.

3.1.2 Prototype Description

The experience prototyping of PeopleLens used an AI prototype that was designed to provide information about people in the immediate social environment for users who are blind or low vision [27]. The prototype utilises a head-mounted augmented reality device in the form of a modified HoloLens [30] in combination with five state-of-the-art computer vision algorithms to *continuously* identify and track people in space as well as capture the gaze direction and activities of people in the vicinity. By tracking the 6DOF motion of the head-mounted user device, information is rendered to users via spatial audio to create a real-time audio augmented reality experience that is responsive to the dynamic flow and movement of social interaction.

The AI prototype employs a pipeline of state-of-the-art computer vision algorithms. The pipeline receives images from four of the sensors on the HoloLens: three grey-scale cameras that provide a near 160-degree field-of-view with good vertical capture and a 1344x756 pixel central RGB colour camera. Images are first processed by a model to detect the pose and location of all people in a scene (re-implementation of [12]). Poses are then used to crop the face, which is passed on to a probabilistic model that provides a set of potential identities [14]. The poses are also fed into models that provide information on gaze direction (implementation of [42]) and activity (unpublished). A responsive experience is achieved through maintaining a predicted world state of the social environment in a four-meter radius for all 360 degrees around the user.

Computation is done on a server box containing two 24 GB TitanX GPUs. While compute heavy, it is reasonably fast, with the world state model updating at 12 to 15 frames per second and the user experience working at 60 frames per second. The experience is provided on a modified HoloLens mixed reality device, with the display (lens) removed (Figure 2). This adaptation reduced the weight of the wearable, avoided occlusion of any residual vision of the user, and provided a more natural experience of communication for conversation partners. This AI system differs substantially from existing technologies, e.g., SeeingAI [47] or Orcam [37], in the immediacy of the *real-time*, *spatial*, and *continuous* experience achieved.

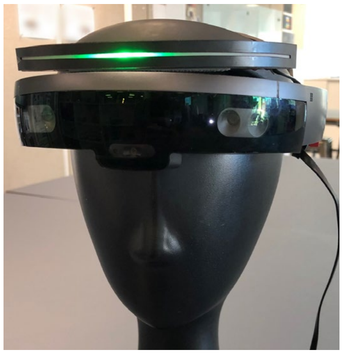

Figure 2: Modified HoloLens device with external LED display.

3.1.3 Baseline Performance

As performance can impact usage, we calculated the baseline performance of the AI prototype on a dataset of children carrying out typical school activities. All performance metrics were calculated on a per-frame basis and averaged over the whole of the dataset. The dataset was formulated alongside the children as part of outreach work in helping children learn about AI systems and was subsequently donated to the project. It contains 14m 08s of footage from 19 recordings of 4 distinct activities, such as eating lunch or entering a classroom, and included 18 children. This dataset was annotated with ground truth labels. The presence of people was correctly identified in the frame with an accuracy of 79% (P: 97% and R: 79%). Per frame correct identification of a named person had an accuracy of 18% (P: 22% and R: 18%). These numbers reflect the ‘raw’ predictions of the AI prototype and do not include performance enhancements achieved through tracking and the interaction design, which are much more difficult to measure.

3.1.4 Data Considerations

Datasets and handling are a significant part of the ethical deployment of any AI system, particularly this one which raises numerous value tensions [19]. This project followed the Responsible AI Principles from Microsoft, assessing potential harms and needed mitigations in the context of the overall benefit of the system [68]. The system was built to avoid storing image, time-stamp or location data that could be misused by a malicious actor. While a number of models were used to support identity recognition, the facial recognition was done with a proprietary model that had been tested on a broad set of demographics. We also considered the impact of the system not recognizing a child with facial anomaly. While we were unable to implement a technical solution in our deployment, we prepared a protocol for addressing the issue quickly should it arise. This included priming teachers to look for inconsistent recognition of specific people as well as a data capture plan to analyse and address the situation.

3.2 Final PeopleLens Experience

In order to provide the context for our discussion of the design journey in Section 4, we describe the final PeopleLens experience that resulted from that journey first.

3.2.1 PeopleLens Features

The PeopleLens has three key features:
Person-In-Front: This feature reads out the name of a person when TH looks at them.

- All sounds are spatialised so that the person’s name is heard from the direction of that person at the time it is read out.
- If TH moves his head quickly, the notification triggers when TH’s gaze crosses the nose of a person, but the sound is spatialised according to head position at the time of rendering.

Orientation Guide: This feature provides additional sound cues to support TH’s in situ understanding of the detection of bodies or faces. These cues assist TH in orienting his body and head to interact in a socially understood way as well as provide good images to the system for processing.

- Spatialised percussive bumps are played when a body is seen.
- Bumps are followed by a name if the person is known to the system and identifiable.

- If a person hasn’t been identified after 0.1seconds, a click sound is played.
- If TH dwells on a click sound for 2 seconds, an elastic band sound is used to help TH shift his gaze up or down to help him orient towards the face.

External Feedback: Finally, the head-mounted device has an external LED interface (see Figure 2) that indicates the system state to a communication partner:

- A white light shows the position of the closest tracked person within the system’s 180° field of vision.
- The white light turns green when the system reads out a name to TH.

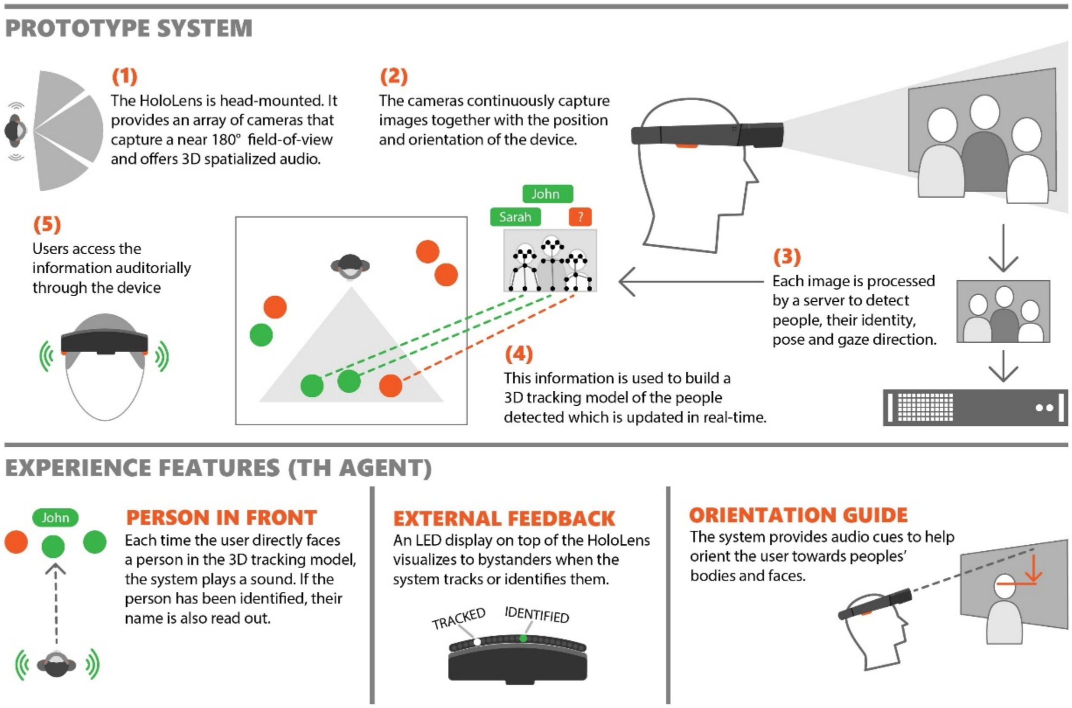
PROTOTYPE SYSTEM
(1)
The HoloLens is head-mounted. It provides an array of cameras that capture a near 180° field-of-view and offers 3D spatialized audio.
(2)
The cameras continuously capture images together with the position and orientation of the device.
(5)
Users access the information auditorially through the device
John
Sarah
?
(3)
Each image is processed by a server to detect people, their identity, pose and gaze direction.
(4)
This information is used to build a 3D tracking model of the people detected which is updated in real-time.
EXPERIENCE FEATURES (TH AGENT)
John
PERSON IN FRONT
Each time the user directly faces a person in the 3D tracking model, the system plays a sound. If the person has been identified, their name is also read out.
EXTERNAL FEEDBACK
An LED display on top of the HoloLens visualizes to bystanders when the system tracks or identifies them.
TRACKED
IDENTIFIED
ORIENTATION GUIDE
The system provides audio cues to help orient the user towards peoples’ bodies and faces.

Figure 3: PeopleLens experience features

3.2.2 Exemplar Scenario of Use

Let’s take the following scenario of use proposed by TH to understand how the PeopleLens might work. This is illustrative of the connected use of the PeopleLens system, however, we chose not to implement this particular scenario to avoid disrupting classroom learning if the technology went wrong.

TH walks into a classroom that he is very familiar with. He hears three “bumps” (percussive sounds representing body presence) at about 10 o’clock (forward left). TH guesses that three people are standing at the interactive whiteboard working on a problem set with their backs to him. As he shifts his gaze to the right, he hears a smattering of bumps which he guesses are other children sitting at their desks reading with their heads down. As he gets all the way to the right, he hears a bump followed by a name, Jane. TH clicks his tongue and listens for its echo. He guesses that Jane is standing next to the wall, perhaps at the classroom coat rack. However, TH really wants to tell his friend Oscar about the new Lego set he got over the weekend. He heads in the direction of Oscar's seat. As he gets closer, he hears woodblock sounds which encourage him to look down. Looking at the external interface on TH's headset, Oscar can see that the system sees him but he's to the right of center so he needs to shift a bit to be properly detected. Oscar's name is then read out. TH surmises that Oscar must be sitting down. TH grabs a chair and pulls out a Lego figure for Oscar to see.

4 PEOPLELENS DESIGN JOURNEY

Now that we have described the "destination" in terms of the final prototype we arrived at, we reflect on the journey of designing the human-AI interaction for the PeopleLens. We put particular emphasis on the ways in which the design and design process address, or contribute to our understanding of, capability uncertainty. To do this, we take two different perspectives on the design journey. The first captures the evolution of key design decisions and the ways they addressed the uncertain outputs of a multi-algorithm system working in the real world. The second takes a higher-level view of how TH's notions of the capabilities of the AI prototype change, shaped by his and others' interactions with it. We preface these two sub-sections with a description of designing with TH and how we documented the journey.

4.1 Introducing TH

Given both the technical and ethical challenges of deploying a multi-algorithm, state-of-the-art AI system that includes facial recognition in a school, we felt the responsible approach [46] was to first focus on a deep engagement with a single young blind person, their family, and their school community. We chose to work with TH, a 12-year-old boy born blind who attends an independent mainstream school. A very able child, he loves mathematics and coding. He accesses all learning through tactile means (e.g., braille, objects) as he has light perception only. Despite being a very independent child with excellent mobility skills, TH finds it difficult to interact with his peers in school. His family is constantly encouraging him to adopt more socially oriented communication strategies, from saying hello when he meets someone to not putting his head on the desk when people are talking to him. He also struggles to engage in extended conversations without mentioning technology, his favorite topic.

Our choice to work with TH was motivated by various practical considerations. He and his family had already worked with our research team, and they were excited by the technology they had glimpsed. As TH had previously been a co-designer on a different four-year project from sketches to commercialisation, he was familiar with technology design processes and understood the possibility that this research technology might never get to product and into his hands. We felt this was an important ethical consideration. Ethical clearance was received through all necessary IRB processes for this co-design work and inclusion of pictures in this paper. Consent was also sought and provided by all participants (or their parents) in the school setting.

4.1.1 Designing with TH

Our seven-month engagement took place in two phases. Design explorations began in our lab with three 1.5-hour-long monthly co-design sessions in which TH participated with his mother. The aim was to use the AI prototype to support TH and his mother in imagining what scenarios the technology might be most useful to TH and how we could refine the experience for these initial scenarios. We scaffolded this imagination process by providing different functionality for TH to try and discuss. It also gave us an opportunity to observe potential design issues and hear from TH about his expectations of the ultimate experience design. Having identified an initial experience and use scenario, we then engaged with TH's school community to explore the experience in the wild. The first two sessions were small steps to explore the technology outside the lab: one in TH's home, and the other in a controlled school setting with three other children. In the final two sessions, TH used the technology during a drama class.

4.1.2 Understanding our engagement with TH

In order to document and reflect on the design process and the capability uncertainties we encountered, data was collected throughout the experience prototyping process. The initial lab sessions, the family meal in TH's home, and the initial school session were audio and video recorded. The video stream and system annotations (e.g., recognised people) from the PeopleLens were also captured. All discussions in the audio recordings were transcribed. System data and annotations were reviewed in custom-written visualisation software.

In the whole-class school environment (final two sessions) no video or audio was collected, neither from an external camera, nor the HoloLens. The user experience was recorded and could be anonymously played back through custom tooling, using 6DOF measures from the HoloLens as well as pose annotations. To support this as a non-visual process, a researcher had an interface for annotating key events and capturing time-stamped observational notes. Observational notes were also taken by an additional researcher. TH and his mother were interviewed after each session. Email feedback was sought from involved teachers. The parents of 23 of the 30 children that the system could potentially identify gave consent for the PeopleLens to learn the children's identities from images. These images were immediately discarded after training.

The methodological perspective taken in the analysis was interpretive. Acknowledging the mixture of data materials produced throughout the research process, we sought to apply different analytical techniques and expand on salient excerpts and descriptions in the data in group research meetings and workshops. Some of the sequences were then subject to interaction analysis, examining in detail the turns of talk and interaction that unfolded around TH's use of the PeopleLens. Interesting features from the data were also algorithmically extracted and visualised in our custom tooling to explore specific patterns of interaction in more detail. Alongside these empirically focused activities, theory from accessibility design and disability studies was used to help develop and thicken the analysis, building the interpretive frame for thinking about our research.

4.2 Design Evolution

4.2.1 Exploration in the Lab

In the first session, TH was offered a number of different experiences in isolation: 1) *person-in-front*, in which the name of a person was read out when TH looked at them; 2) *gaze upon*, a spatialised cue when someone looked at TH; and 3) *follow me*, an ongoing sound that tracked the person in closest communication distance. For each experience, TH was asked to walk through his previous day and identify if and when such functionality would have been useful. This process was supported by his mother using appropriate tactile materials known to TH. TH generated a large number of use scenarios, but his overwhelming focus was on his desire to have all three features in one experience. He asked many times, "Can I have it all at the same time?" This became the focus of the second session.

In the second session, TH was given a version of the AI prototype that allowed him access to all of its functionality through a wrist-worn wearable. In addition to the functionality above, he could also access the number of people looking at him and an overview of all tracked people. However, we found that he faced a constant tension between wanting more information about his environment and the practical constraint that he had to spend significant energy processing the information he received. This was particularly true for information he did not typically account for in his interactions, such as social information (e.g. who is around him). By the end of the session, TH flopped on the floor, exhausted.

As a result, **low information density** became a deliberate design decision, so that TH could focus on the information most salient in any interaction: who is there. Despite the wide range of capabilities available in the AI Prototype, the PeopleLens experience is configured to read out a name only when TH looks directly at someone (Figure 3: Person-in-front).

In the final lab session, TH was given the Person-in-front experience to try. We soon discovered that TH needed additional **audio cues to support orientation and dynamic intelligibility** to make a low information density interface useful. We had to consider how the dynamic experience of moving through space to find people could be supported. We developed two types of sound cues to aid TH's awareness of, and orientation towards, other people. The system plays a "bump" – a short percussive sound – when it identifies people's bodies within its field of view. These bump sounds are rendered using spatial audio, providing TH fast feedback on the position of other people regardless of their distance or orientation. The speed with which these bumps can be produced (as opposed to names) allows TH to better direct his attention towards a person (or people) in space.

If the bump sound is not followed by a name, a natural "woodblocks" sound is played to help orient TH to the nearest face. The woodblock sound changes in pitch if TH's head is tilted too high or too low, with a clear snap sound when TH centres on a face. While the bump sound serves to assist TH in locating people's bodies, the woodblock sound aids in locating people's faces. In having dynamic access to these kinds of information, TH can work with the AI system to support its ability to detect and identify people more robustly, regardless of lag, image quality, or person orientation (Figure 3: Orientation Guide).

Having tried the system, TH was then asked to ideate on which scenario of use he was keen to try first. As trying this out in a noisy, 100+ person lunch hall (TH's first choice) was not technically feasible, TH decided on a busy classroom environment. Classrooms busy with children are often noisy, which made it particularly difficult for TH to rely on audio cues to navigate social situations.

4.2.2 Explorations in the Wild

With an initial scenario of use and design in place, the final four sessions took place outside the lab. First, the research team joined TH and his extended family for a big Sunday lunch with seven adults present. The PeopleLens was used first as TH moved from room to room to find people and was then used again during lunch with everyone present around a dining table. It was worn for nearly an hour. In the second session, which lasted 45 minutes, the PeopleLens was used with classmates. TH played a search-and-rescue game in which he had to find his classmates and bring them to a safe zone. PeopleLens was then used during a Lego activity challenge in which the team had to build the longest possible bridge from Lego. The session finished with the other children trying the PeopleLens blindfolded and speaking about the experience.

The final two sessions took place as part of TH's drama class. The first drama class was in the (very) large school gym. TH was first asked to deliver props to particular classmates. He was then asked to make a map of where the other children were using Lego. Finally, he was left to explore, joining in with the activity if he wanted to. TH did not normally join in for drama due to the interactional challenges of being with 20 children in a space that was difficult for him to navigate due to a lack of furniture. The second drama class, although in a different large open space, took exactly the same format. TH took part in both consecutive sections of his class, a total time of one hour. Both classes were taught by TH's form tutor (homeroom teacher) who knew TH well.

Moving out of the lab encouraged us to focus more strongly on the social situation in which the PeopleLens was used. We realised that the PeopleLens was not just a resource for TH, but also a means of supporting reciprocal interaction with communication partners. We noted that people would often adjust their own body position in order to be identified. Yet, with no feedback from TH or the PeopleLens, it was difficult for a communication partner to know how they might activate this capacity of "being seen." Observations like this eventually led to the development of **external feedback to facilitate reciprocal interaction**. This was intended to assist the development of common ground and reflexive interpretation of behavior, so communication partners could realistically base their interaction on their understanding of what TH knows.

We included an external interface on the HoloLens to support mutual orientation with a communication partner. We respected that TH felt awkward about the idea that others could hear the audio feedback of the PeopleLens. Thus, we designed a semi-circular LED interface affixed to the top of the HoloLens (Figure 2). The display provides communication partners with information about the state of the system. A moving white light tracks the location of the nearest detected person, flashing green when that person is identified to TH. In this way we used different forms of spatial information (sound for TH and light for communication partner) to ensure a shared, reciprocal interaction to which both TH and communication partner could orient (Figure 3: External Feedback).

First deployed in the second of the four sessions, it initially tracked everyone (not just the nearest communication partner). However, the display made the device resemble Christmas tree lights and was therefore insufficient in supporting mutual orientation because people could not identify which specific light represented themselves. As those in the distance are less likely to use the information provided by the external display, we focused on providing feedback to the closest communication partner. The final design was used in the drama classes. In this limited circumstance, we observed how children and adults use this feature to be seen, and occasionally to hide, in interaction with TH, suggesting the display is filling its purpose in mutually negotiating interaction.

4.3 System Capability Evolution

Having looked at design decisions in the previous section, we now take a higher-level view to consider how TH's notions of the capabilities of the AI system changed throughout the design process and usage.

4.3.1 Imagination to Usage

During the design process, TH articulated that the main intended purpose of the PeopleLens should be as a tool to help him find his friends at school. Despite TH's expressed desire in the design process, we saw that he used it to monitor the whereabouts of his teachers instead. In the first school session, TH sat around a square table with three other children tasked with collaboratively building the longest Lego bridge possible. TH's two teaching assistants, his class teacher, the head of special educational provision, his mother, and the researchers were scattered around the edges of a large classroom. TH did not use the PeopleLens to see where his classmates were and did not look at them when he spoke with them. He did, however, call out the names and greet his various teachers as they moved about and came and went from the classroom. He was particularly interested in how he could use the PeopleLens to figure out who had left the room when he heard the door open.

Upon first usage, we see that TH's notion of the capability of the PeopleLens shifts from information about friends to adults. While this behavior was not what the research team expected, it did not surprise TH's mother. She commented: "Knowing where the adults are keeps you from getting in trouble." To give context, TH often felt he was being watched, because he continually got "told off." The view from TH's mother is that TH was unaware who was around and therefore TH had a more difficult time adjusting his behavior to the context. As has previously been pointed out [33], the assistance bubble that surrounds blind children in mainstream schools can make it difficult to 'play footsie under the table'. While TH and his mother had bigger aspirations during the design process for improving social interaction, TH first used the PeopleLens to extend his existing social orientation strategies (locating his teachers) to address his priorities when it came to real social experiences in school.

It is with this unexpected use of PeopleLens that we can begin to develop and refine a notion of capability. Yang et al. [64] draw attention to the uncertain capabilities of AI systems due to data and environment interaction. We see from the above that TH is also very much a part of this emerging set of capabilities. With the PeopleLens, he orients himself to others in ways that, while unanticipated, provide him with a very different ability to engage in conversation. It was in working with TH and the PeopleLens that we began to understand capability in a broader sense, as not just something that develops in the technical system but in and through the interaction between the user and the AI system.

4.3.2 Physical and Attentional Orientation

What became especially apparent was how the interweaved capabilities of TH and PeopleLens evolved over time. One of TH's most noticeable uses of the PeopleLens was to physically orient towards social interaction. To illustrate this point, we compare two vignettes. One from the first session before TH engaged with the PeopleLens, and one while he was using it during a family meal. In each vignette, we consider the ways TH orients to a social situation and how others respond to that orientation.

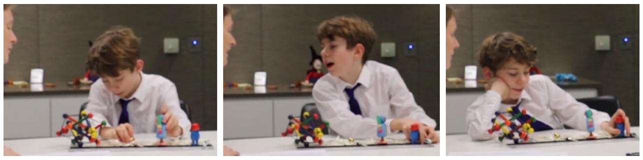

Figure 4. Vignette 1 images

Vignette 1: First Design Session (without PeopleLens). TH is listening to a researcher tell him about the activities for the session. Throughout most of the conversation he is leaning down with his head slightly angled toward the researcher who is speaking, while his hands are busy engaging with the tactile objects of reference for the session (Figure 4: image 1). At one point he lifts his head to the researcher to inject a comment: "I don't want this project to be limited to spaces with super-fast Wifi" (Figure 4: image 2). As the researcher attempts to respond, he interrupts again with a comment on the difficulties of finding super-fast Wifi in public places. He then adopts a posture with his head in his hand as the

conversation continues (Figure 4: image 3). TH's mother uses a variety of tactile cues to encourage TH to stay on topic and adjust his posture to visually indicate his social engagement.

In this first Vignette, TH did little to physically orient to the interaction despite attending to the topic. This suggested to the researchers unfamiliar with TH that he was disengaged from the activity. At one point, he did shift his head towards the researcher who was speaking, but because he did not turn or straighten up his body he gave the impression that his interest was fleeting (and to some extent it was, as TH's interjections were only tangentially related to the topic of conversation). TH's physical posture, signaling disengagement, persisted in many of the interactions the research team had with him despite frequent tactile and verbal reminders from his mother to sit up straight, keep his head up, and turn towards people he was speaking to.

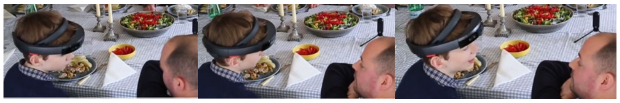

Figure 5. Vignette 2 images

Vignette 2: Family Meal (with PeopleLens): TH is learning about a new feature that the team engineer, Philip, is about to turn on. TH turns towards Philip as he hears him approach and listens (Figure 5: image 1). After asking a question, he shifts his gaze slightly away (Figure 5: image 2) as he attends to the answer. TH then turns towards Philip to ask a follow up question to what Philip has just said (Figure 6: image 3).

In this second vignette, TH communicated active participation in the conversation through his body orientation. He turned his body towards Philip (a pseudonym for this paper) and looked towards Philip's eyes when re-engaging Philip during TH's turn in the conversation. We also noticed that TH was able to deepen the conversation through asking follow-up questions rather than interjecting new topics. This led to a smoother conversation. The difference in interaction was noted by his family in a number of ways:

> "Before it was the ear that was talking to me, now it's his whole person. It's so different." -- TH's grandmother
>
> "He will turn, find them, and kind of look at them, and then he will try to interject. Before he would just say their name and hope for the best." -- TH's father

The contrast between the two vignettes above draws further attention to the way a fluid capability is achieved between TH and the system. With PeopleLens, TH begins to hold his body differently and to participate in conversations in ways that were not available to him before. Despite TH's challenge in articulating how he used the PeopleLens or why he never wanted to give it back, these achieved capabilities were noteworthy to his family. A particular example of TH's behaviour using the system presents a helpful illustration of this adaptive behavior and emerging capability. We observed that TH started to subtly shift his head side-to-side to repeatedly have PeopleLens read out the name of the person(s) with whom he was speaking. At his family meal for example, TH used the PeopleLens for 54 minutes from which we automatically extracted 28 instances of this lateral head shaking. We see similar numbers in the final school sessions (Drama Class 1: 24 minutes, 14 occurrences; Drama Class 2: 23 minutes, 16 occurrences). This struck the research team as odd, as his existing strategies should have easily sufficed to know with whom he was speaking when in a one-to-one conversation.

Visually inspecting the data and accounting for literature on blind children and social interaction, we propose a number of hypotheses for why TH's usage may have developed in this way. We saw, for example, that he might do this when trying, and succeeding, to get someone's direct eye contact. We also observed this process when there were multiple people in close proximity and he repeatedly moved his focus between them. However, as our visualisation data suggested that it happened every 2 minutes or so (the span of working memory), it may also be that TH had discovered he could support his spatial attention by using the system to stabilise and refresh a representation that effectively links identities and space in his working memory. This is necessary to maintain attention to external happenings, which is a significant challenge for many blind children but critical for maintaining a topic of conversation.

While understanding the PeopleLens's usage in this way requires further exploration, it re-emphasises the active process of configuring information resources and illustrates how difficult it might be to predict how such information might be employed to extend capabilities without having prototypes to hand. Interestingly, TH was unaware of this specific behaviour and found it difficult to describe what he wanted to do with the technology in general. This is quite typical of skilled behaviour: habitual or routine actions are often unavailable to us in any conscious or explicit way, as is well known in psychology (see, for example, Kahneman's writings on "Fast and Slow Thinking" [24]).

We found it remains important to recognise the varied ways capabilities take hold and come to have value. For example, even when using the PeopleLens, TH still found it difficult to orient to unstructured social environments, such as his drama class. TH wandered through the students spread out in a large gym, walking through the middle of groups and ignoring people not of interest despite identification by the PeopleLens. We speculate that these different responses might be a function of the increased complexity of the situation. Engaging in his drama class required understanding visual social concepts, e.g., groups, as well as nuanced (and potentially visual) social skills such as conversation initiation. Participating in a family meal with familiar adults allowed TH to focus on single interactional mechanism, e.g., making eye contact when speaking. These differing use patterns suggest that understanding social interaction is not just about knowing where people are, but also incorporating higher order concepts of social interaction that are mainly understood visually (e.g., proxemics), a process that needs to be actively supported as TH learns to understand them. As we will suggest next, it is also clear that the other people in a setting can play an important role in a broader view of capability.

4.3.3 Socially Situated Sensemaking

We found capability to be as much about the interactions between TH and the PeopleLens as it was about the system's inherent functionality. TH and PeopleLens can, if you will, to be capable *together*. Our detailed studies of TH's interactions with his family while using the PeopleLens also showed these interactions extended to include others present in a setting. For example, in one scene from the recorded meal TH confirms that it is his grandmother standing in front of him. He then announces: "Now that's Philip," shifting his head to focus on Philip who has come up behind his grandmother. As Philip is speaking with someone else at the table, it is likely that TH has heard Philip's voice and is on some level aware that Philip is there. However, we might surmise that the announcement of the name is a verbalisation of TH's process to bring PeopleLens recognition and his own sensemaking of the situation together.

The above example illustrates how TH uses the PeopleLens alongside his existing sensemaking strategies to build up a sense of where people are around him. The PeopleLens is acting in concert with other information available to TH, illustrating the ongoing back-and-forth between TH's sensemaking and information from the PeopleLens, and ultimately a highly dynamic and fluid set of capabilities. Critically, the PeopleLens is not a replacement for TH's sensemaking, but instead provides the conditions for different and sometimes new capabilities to emerge. The way TH builds and maintains spatial awareness using the system is key, as it serves to orient him to where the talk is being conducted and how to participate in the conversation. These observations emphasise how a dynamic and adaptive system (like PeopleLens) can be utilised to enable collective sensemaking in ways mundanely achieved, but perhaps not easily imagined.

5 DISCUSSION

This paper presents and reflects upon the user experience prototyping phase of PeopleLens, an open-ended AI system intended to support the social agency of blind children at school, which was conducted with TH, a 12-year-old blind boy. It is an exploration of how we design human-AI interaction that moves beyond discrete task support to provide continuous, dynamic information streams. We share the artefact of this process, the PeopleLens, and vignettes of its use during the prototyping phase as a means to reflect upon designing with capability uncertainty in the creation of open-ended AI experiences. First, we consider the PeopleLens as an exemplar of how AI systems may enable people to extend their own capabilities. The mutual shaping of user, technology, and social context is then used to ground our discussion of key design decisions to enable the human-AI interaction in these kinds of systems. We close by considering how this human-AI interaction evolves notions of system capabilities and purpose.

5.1 Open-ended AI and Extending Capabilities

The PeopleLens artefact and descriptions of its use by TH offer a concrete example of how an open-ended AI system might enable and extend human capability in ways different from the discrete tasks AI systems currently support well. Demonstrated in the prototyping journey, the continuous nature of the system instigated a mutual and ongoing negotiation between TH and PeopleLens – user and AI system. Through the various vignettes of use, we showed how TH used the PeopleLens, and the simple, continuous cues that it provided, as a resource to manage interaction in ways that gave him new capabilities: physically and attentionally orienting to interaction in ways that demonstrated active participation and led to smoother social communication.

The human-AI interaction was one in which the capabilities of each created something new together: the notion of cognitive extension as opposed to cognition as a service [22]. This cognitive extension recognises that the system is not something that acts on behalf of its user or offers a non-existent capability to a user. Rather, interaction with the system adds to the mix of what is already possible: helping a user to achieve their existing goals and aspirations, adding and enriching information already relied upon, and building new strategies on top of existing ones. This more complex and unfolding relationship in turn changes a user's ability to manage complex human processes. This concrete example demonstrates the potential opportunities that an AI experience can provide when working in concert with human agency as a part of the sensemaking process for people with disabilities [5,33].

While the AI experience we designed was specific to the user (TH), we also highlighted the importance of supporting socially situated interaction. We saw that shared attention was a complex achievement, particularly with different sensory modalities at play. The PeopleLens appeared to offer some resolution, not in that it offered TH a way to "see," but rather it provided him with a means of mutually constructing a frame of reference with others. Through his voicing of the PeopleLens's output and the more active relations he built with others using turn-taking and head movements, TH was able to construct a shared understanding. The ensemble of joint, continuous interactions between TH, others in his proximity, and the PeopleLens enabled the establishment of new forms of social communication.

While assistive technologies are often designed with a narrow focus on the interactions between user and system, a shift in emphasis towards extending human capabilities will demand a much greater sensitivity to a wider set of social relationships. A key point to recognise here is that the meanings and actions available in a setting are actively produced by the actors involved, including the AI system (connecting to Suchman's seminal work [55]). To extend human capabilities, then, is to consider how design might aid such mutual productions of meaning and action, and do so in ways that are accessible to diverse users. A further point to consider is how such meanings and actions play into long established norms, and in some cases reproduce ableist norms that sanction the objectification and discrimination against people with disabilities [6]. In the work we present, for example, further research is needed to examine how mutual attention might be signified not just through bodily orientations that arguably align with sighted norms, but also in more diverse ways that expand the repertoire for interaction.

5.2 Designing Intelligible Open-ended AI

Building an intelligible, long-term relationship with the system also meant making deliberate design choices so that the system was easily comprehensible to both the primary user of the technology and other people nearby. We identified three key aspects of information design to support social sensemaking: 1) *low information density* enables the user to remain focused on the world around them and develop their own appropriation strategies in social situations; 2) *dynamic intelligibility* provides in situ sound cues that guide the user's body and head orientation towards people and faces. This dynamic interaction with the PeopleLens helps the user find people and demonstrate the user's communication intent, while yielding more robust people identification; and 3) the addition of *external feedback* that visually communicates the system state to bystanders invites a level of intelligibility that enables them to mutually orient to the user of the system. Together, we believe these design considerations enable users to adapt their use of the system as their experience and the context of use changes.

The relative simplicity of the features offered by the PeopleLens experience compared to the opportunity provided by the AI prototype demonstrates that extending capabilities is not necessarily about providing functional complexity but thoughtful design. Despite the technical sophistication of the AI system, what supported TH best was the provision of simple, well-chosen cues as a set of resources at his disposal. Both designers and users may be seduced by the desire for as rich an information experience as possible, and as technology develops, this can lead to a surfeit of information. However, low information density becomes necessary when the AI system becomes embedded in everyday life and users need and want to engage with the world around them. The challenge is not a straightforward matter of too much or too little, but how information in a user's surroundings can be provided such that it is a resource for capabilities to be extended [54].

The mutual shaping of user and technology also required a different approach to *intelligibility*, or the user's understanding of the AI system. Multiple audio cues were used to guide the user towards faces, an interaction that indicated communication intent for the user but also improved the ability of the PeopleLens to recognise people. The more a user works with these cues, the better experience they can gain from the system. Intelligibility of AI has often focused on explaining an outcome [17,32]. In an open-ended system in which there is give-and-take between user and AI system, we would suggest that intelligibility comes from understanding how to improve the outcome of the system, not necessarily understanding all of the technical workings of it. As such, we may need to think about how future research in intelligibility applies to open-ended systems.

The mutual shaping of user and technology can also be socially situated and shaped, as we saw through the documentation of PeopleLens's usage by TH. As designers, we must think about the reciprocal interactions with others in a setting and how such interactions are supported. In the case of the PeopleLens, one key design element is the external-facing interface, giving those engaging with TH a means to see the information the PeopleLens was providing. We saw that this was key to both parties mutually orienting to the interaction. It also gave the communication partner a way to actively "be seen" by TH, helping both to initiate and adapt interaction. Intelligibility of AI systems for use in social contexts may need to account for a wider set of inspections during everyday use, beyond the user. We must explicitly consider how we are designing to account for bystander needs, e.g., their ability to orient to the AI system.

The very types of mutually shaped interactions that intelligible design enables also raises new challenges in measuring success. Metrics are a key part of tuning AI systems and their definition is key to systems' ultimate capabilities. In this paper, we have presented metrics of recognition speed and accuracy on a per-frame basis. While these "bench" metrics provide important information about system performance, they ignore what happens when users actively collaborate with the system to improve accuracy. Users may adapt to the idiosyncrasies of a system or identify contexts and behaviors that improve its utility in ways that are very difficult to measure. This highlights a challenge that is likely to be shared across open-ended AI systems: how to measure success for evaluation purposes as well as AI system optimisation.

5.3 Ideating Open-ended AI Capabilities

In their review paper, Yang et al. [64] identify one of the key challenges for designers of AI systems as the difficulty in purposefully using AI. This is underpinned by challenges in understanding what the AI can do in situ. Our example shows that users have similar challenges. TH and his mother had little difficulty articulating scenarios in which TH needed more social information, but these differed very much from how TH actually used the PeopleLens. This discrepancy between imagined and actual use may come from the fact that, as noted in the disability literature, AI systems are not a replacement for a "missing" capability. Rather, such systems are part of an assemblage of resources that TH can actively coordinate to extend his own capabilities. Focusing the design process of open-ended AI systems around a lack of capability [35], the need for a superpower [33], or inaccessible tasks [52] will not address this initial capability uncertainty.

We have found, instead, that designing AI systems that serve as a resource for extending capabilities demands sensitivity to the ways that any intervention plays out in the real world. In considering the design process, a working AI prototype was central to involving users in complex discussions of what AI could and could not do. It further took the ideation beyond discussion to specifying more concrete ways that digital and physical resources might be combined to enable new capabilities through a process of appropriation. That said, developing cutting-edge technology is an investment and challenge in its own right. The proposition of a new step in the design process moves us closer to acknowledging the need to understand AI system capabilities in situ [64], but does not address the resource requirements.

In response to the issue of resource demand for deploying a working AI prototype, we would also suggest that tools that help visualise and make tangible how AI prototypes and systems are being used could become a new, standard design tool. For example, in the custom tooling that we built, we could visually align 3D representations of the space with camera streams and recognition events produced by this sensory and multi-model system. We could then search across this data to visualise events highlighted by researchers or participants as we did when inspecting TH's use of the PeopleLens for attentional support. Such tools might contribute to the prototyping phase, but they may also blur the boundaries between prototyping and deployment. Similar to A/B testing on the web [51], the difference becomes one of scale. Thoughtful approaches to scale may also be one of the best ways to assess and address potential harms of the AI system that might surface when used in situ [46].

5.4 Limitations

This work focused on the experience prototyping stage of a much longer journey toward creating open-ended AI experiences to enable people to extend their own capabilities. As a result, the description of the development and details of the AI prototype is short. This work also does not extend to a full evaluation of the system, which is subsequent work. Not least, the focus on designing human-AI interaction has meant that we've only indirectly communicated our stance on the development of assistive technologies and the specifics of working with people who are blind and low vision. While our team has put much thought into all these aspects of the work, for greatest clarity, we have chosen to focus specifically on capability uncertainty in this paper. Any substantive consideration of responsible innovation would require a great deal of nuanced discussion, so we are planning a future publication to address in detail our reflections on creating AI systems with and for people with disabilities, addressing issues such as stigma [48], bias [60], inclusion [65], and privacy [3].

6 CONCLUSION

The nuance of human-AI interaction has been evolving for some years now, providing an alternative to thinking about AI as something that competes against or replaces people. Research and the productisation of technologies demonstrate how AI can provide task-based support for the general public as well as people who are blind or low vision. This paper looks toward the future when AI technologies will allow for more continuous experiences, exploring what human-AI interaction in these open-ended AI systems might look like. To contribute to an imagination of this future and how it might enable AI to extend human capability more work is needed to deploy and reflect upon AI systems in the real world.

ACKNOWLEDGEMENTS

We would like to acknowledge the participants in this study and all formative studies.

REFERENCES

1. Ashraf Abdul, Jo Vermeulen, Danding Wang, Brian Y. Lim, and Mohan Kankanhalli. 2018. Trends and trajectories for explainable, accountable and intelligible systems: An HCI research agenda. In *Proceedings of the 2018 CHI Conference on Human Factors in*

Computing, 1–12.

2. Dragan Ahmetovic, Roberto Manduchi, James M. Coughlan, and Sergio Mascetti. 2015. Zebra Crossing Spotter: Automatic Population of Spatial Databases for Increased Safety of Blind Travelers. 251–258.

3. Taslima Akter, Tousif Ahmed, Apu Kapadia, and Swami Manohar Swaminathan. 2020. Privacy Considerations of the Visually Impaired with Camera Based Assistive Technologies: Misrepresentation, Impropriety, and Fairness. In *The Proceedings of the 2020 SIGACCESS Conference on Computers and Accessibility*.

4. Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Samsi Iqbal, Paul Bennett, Inkpen Kori, and Jaime Teevan. 2019. Guidelines for Human-AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–13.

5. Cynthia L. Bennett, Erin Brady, and Stacy M. Branham. 2018. Interdependence as a frame for assistive technology research and design. In *Proceedings of the 2018 SIGACCESS Conference on Computers and Accessibility*, 161–173.

6. Cynthia L. Bennett, Daniela K. Rosner, and Alex S. Taylor. 2020. The Care Work of Access. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–15.

7. Jeffrey P. Bigham, Tom Yeh, Chandrika Jayant, Hanjie Ji, Greg Little, Andrew Miller, Robert C. Miller, Robin Mille, Aubrey Tatarowicz, Brandyn White, and Samuel White. 2010. VizWiz: nearly real-time answers to visual questions. In *Proceedings of the 2010 UIST symposium on User Interface Software and Technology*, 333–342.

8. Jeffrey P Bigham and Patrick Carrington. 2018. Learning from the Front: People with Disabilities as Early Adopters of AI. In Proceedings of the 2018 *HCIC Human-Computer Interaction Consortium*.

9. Jeffrey R Blum, Mathieu Bouchard, and Jeremy R. Cooperstock. 2013. Spatialized audio environmental awareness for blind users with a smartphone. *Mobile Networks and Applications* 18, 3: 295–309.

10. C. J. Cai, S. Winter, D. Steiner, L. Wilcox, and M. Terry. 2019. “Hello AI”: Uncovering the Onboarding Needs of Medical Practitioners for Human-AI Collaborative Decision-Making. In *Proceedings of the CSCW conferennce on Human-computer Interaction, 3*, 1–24.

11. Carrie J. Cai, Emily Reif, Narayan Hegde, Jason Hipp, Been Kim, Daniel Smilkov, Martin Wattenberg, Fernanda Viegas, Greg S Corrado, Martin C Stumpe, and Michael Terry. 2019. Human-centered tools for coping with imperfect algorithms during medical decision-making. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–14.

12. Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. 2017. Realtime multi-person 2d pose estimation using part affinity fields. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 7291–7299.

13. Giulia Cappagli, Sara Finocchietti, Elena Cocchi, Giuseppina Giammari, Roberta Zumiani, Anna Vera Cuppone, Gabriel Baud-Bovy, and Monica Gori. 2019. Audio motor training improves mobility and spatial cognition in visually impaired children. *Scientific reports* 9, 1: 1–9.

14. Daniel Coelho de Castro and Sebastian Nowozin. 2018. From Face Recognition to Models of Identity: A Bayesian Approach to Learning about Unknown Identities from Unsupervised Data. In *European Conference on Computer Vision (ECCV)*, 745–761.

15. Sylvain Ferrand, Francois Alouges, and Matthieu Aussal. 2018. An augmented reality audio device helping blind people navigation. *International Conference on Computers Helping People with Special Needs*: 28–35.

16. Alexander Fiannaca, Ilias Apostolopoulous, and Eelke Folmer. 2014. Headlock: a wearable navigation aid that helps blind cane users traverse large open spaces. In *Proceedings of the 2014 ASSETS Conference on Computers & Accessibility*, 19–26.

17. Forough Poursabzi-Sangdeh, Daniel G. Goldstein, Jake M. Hofman, Jennifer Wortman Vaughan, and Hanna Wallach. 2018. Manipulating and measuring model interpretability. *arXiv preprint* arXiv:1802.

18. Euan Freeman, Graham Wilson, Stephen Brewster, Gabriel Baud-Bovy, Charlotte Magnusson, and Hector Caltenco. 2017. Audible beacons and wearables in schools: Helping young visually impaired children play and move independently. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 4146–4157.

19. Batya Friedman and David G. Hendry. 2019. *Value sensitive design: Shaping technology with moral imagination*. MIT Press.

20. Linda Hagood. 2008. *Better Together: Building relationships with people who have visual impairment and autism spectrum disorder (or atypical social development).*

21. Jeffrey Heer. 2019. Agency plus automation: Designing artificial intelligence into interactive systems. *Proceedings of the National Academy of Sciences* 116, 6: 1844–1850.

22. José Hernández-Orallo and Karina Vold. 2019. AI extenders: The ethical and societal implications of humans cognitively extended by AI. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 507–513.

23. Hernisa Kacorri, Kris M. Kitani, Jeffrey P. Bigham, and Chieko Asakawa. 2017. People with Visual Impairment Training Personal Object Recognizers: Feasibility and Challenges. In *Proceedings of the 2017 CHI Conference on Human Factors in Computer.*

24. D Kahneman. 2011. *Thinking, fast and slow*. Macmillan.

25. Christine Kelly. 2013. Building bridges with accessible care: Disability studies, feminist care scholarship, and beyond. *Hypatia* 28, 4: 784–800.

26. Marcella Mandanici and Antonio Rodà. 2020. *Large-Scale Interactive Environments for Mobility Training and Experience Sharing of Blind Children*. In Technological Trends in Improved Mobility of the Visually Impaired (pp. 301-318). Springer.

27. Cecily Morrison, Martin Grayson, Anja Thieme, Rita Marques, Daniela Massiceti, Edward Cutrell. 2020. A Dynamic AI System for Extending the Capabilities of Blind People. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems Extended Abstracts*.

28. Laura Mauldin. 2014. Precarious plasticity: Neuropolitics, cochlear implants, and the redefinition of deafness. *Science, Technology, & Human Values* 39, 1: 130–153.

29. Oussama Metatla and Clare Cullen. 2018. "Bursting the Assistance Bubble": Designing Inclusive Technology with Children with Mixed Visual Abilities. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 346.

30. Microsoft Hololens. *Https://www.microsoft.com/en-us/hololens*.

31. Jennie Middleton and Hari Byles. 2019. Interdependent temporalities and the everyday mobilities of visually impaired young people. *Geoforum* 102: 76–85.

32. Grégoire Montavon, Wojciech Samek, and Klaus-Robert Müller. 2018. Methods for interpreting and understanding deep neural networks. *Digital Signal Processing* 73: 1–15.

33. Cecily Morrison, Edward Cutrell, Anupama Dhareshwar, Kevin Doherty, Anja Thieme, and Alex Taylor. 2017. Imagining Artificial Intelligence Applications with People with Visual Disabilities using Tactile Ideation , pp. 81-90. ACM, 2017. In *Proceedings of the 2017 SIGACCESS Conference on Computers and Accessibility*, 81.

34. Ingunn Moser. 2006. Disability and the promises of technology: Technology, subjectivity and embodiment within an order of the normal. *Information, Communication & Society* 9, 3: 373–395.

35. Lauren Murray, Philip Hands, Ross Goucher, and Juan Ye. 2016. Capturing social cues with imaging glasses. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 968–972.

36. Changhoon Oh, Jungwoo Song, Jinhan Choi, Seonghyeon Kim, Sungwoo Lee, and Bongwon Suh. 2018. I lead, you help but only with enough details: Understanding user experience of co-creation with artificial intelligence. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–13.

37. Orcam My Eye. https://www.orcam.com/en/.

38. Sethuraman Panchanathan, Shayok Chakraborty, and Troy McDaniel. 2016. Social Interaction Assistant: A Person-Centered Approach to Enrich Social Interactions for Individuals With Visual Impairments. *IEEE Journal of Selected Topics in Signal Processing* 10, 5.

39. Sabrina A. Paneels, Dylan Varenne, Jeffrey R. Blum, and Jeremy R. Cooperstock. 2013. The walking straight mobile application:

Helping the visually impaired avoid veering. In Proceedings of the 2013 ICAD conference on *International Conference on Auditory Display, 25 -32*.

40. Peter Pirolli and Daniel M. Russell. 2011. Introduction to this special issue on sensemaking. *Human-Computer Interaction* 26, 1–2.

41. Linda Pring (Ed.). 2005. *Autism and blindness: Research and reflections.*

42. Sergey Prokudin, Peter Gehler, and Sebastian Nowozin. 2018. Deep directional statistics: Pose estimation with uncertainty quantification. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 534–551.

43. Shi Qiu, Matthias Rauterberg, and Jun Hu. 2016. Tactile Band: Accessing Gaze Signals from the Sighted in Face-to-Face Communication. In *Proceedings of the 2016 TEI Conference on Tangible, Embedded, and Embodied Interaction*, 556–562.

44. Daniel M. Russell, Mark J. Stefik, Peter Pirolli, and Stuart K. Card. 1993. The cost structure of sensemaking. In *Proceedings of the the 1993 CHI Conference on Human Factors in Computing Systems*, 269–276.

45. Sharon Sacks and Robert Gaylord-Ross. 1989. Peer-mediated and teacher-directed social skills training for visually impaired students. *Behavior Therapy* 20, 4: 619–640.

46. Daniel Schiff, Bogdana Rakova, Aladdin Ayesh, Anat Fanti, and Michael Lennon. 2020. Principles to Practices for Responsible AI: Closing the Gap. *arXiv preprint*: https://arxiv.org/abs/2006.04707.

47. SeeingAI. *Https://www.microsoft.com/en-us/ai/seeing-ai*.

48. Kristen Shinohara and Jacob O. Wobbrock. 2011. In the shadow of misperception. *Proceedings of the 2011 CHI Conference on Human Factors in Computing Systems,* 705.

49. Ben Shneiderman. 2020. Design Lessons From AI's Two Grand Goals: Human Emulation and Useful Applications. *IEEE Transactions on Technology and Society* 1, 2: 73–82.

50. Ben Shneiderman. 2020. Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy. *International Journal of Human–Computer Interaction* 36, 6: 495–504.

51. Dan Siroker and Pete Koomen. *A/B testing: The most powerful way to turn clicks into customers*. John Wiley & Sons.

52. Abigale J. Stangl, Esha Kothari, Suyog D. Jain, Tom Yeh, Kristen Grauman, and Danna Gurari. 2018. BrowseWithMe: An Online Clothes Shopping Assistant for People with Visual Impairments. In Proceedings of the *2018 SIGACCESS Conference on Computers and Accessibility*, 107–118.

53. Lee Stearns, Ruofei Du, Uran Oh, Yumeng Wang, Leah Findlater, Rama Chellappa, and Jon E. Froehlich. 2014. The design and preliminary evaluation of a finger-mounted camera and feedback system to enable reading of printed text for the blind. In *European Conference on Computer Vision*, 615–631.

54. Lee Stearns, Leah Findlater, and Jon E. Froehlich. 2018. Applying Transfer Learning to Recognize Clothing Patterns Using a Finger-Mounted Camera. In *Proceedings of the 2018 SIGACCESS Conference on Computers and Accessibility*, 349–351.

55. Lucy A Suchman. 1987. *Plans and situated actions: The problem of human-machine communication.* Cambridge University Press.

56. Valerie Tadić, Linda Pring, and Naomi Dale. 2010. Are language and social communication intact in children with congenital visual impairment at school age? *Journal of Child Psychology and Psychiatry* 51, 6: 696–705.

57. Anja Thieme, Cynthia L. Bennett, Cecily Morrison, Edward Cutrell, and Alex S. Taylor. 2018. "I can do everything but see!--How People with Vision Impairments Negotiate their Abilities in Social Contexts. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 203.

58. Suzanne H Verver, Mathijs PJ Vervloed, and Bert Steenbergen. 2019. The use of augmented toys to facilitate play in school-aged children with visual impairments. *Research in developmental disabilities* 85: 70–81.

59. Karl Weick. 1995. *Sensemaking in Organisations*. Sage, London.

60. Linda Wong and Alexander Wang. 2019. Implications of Computer Vision Driven Assistive Technologies Towards Individuals with Visual Impairment. *arxiv* arXiv:1905.

61. Shaomei Wu, Jeffrey Wieland, Omid Farivar, and Julie Schiller. 2017. Automatic Alt-text: Computer-generated Image Descriptions for Blind Users on a Social Networking Service. In *Proeceedings of the 2017 CSCW Conference on Computer-Supported Cooperative Work and Social Computing*.

62. Qian Yang, Justin Cranshaw, Saleema Amershi, Shamsi T. Iqbal, and Jaime Teevan. 2019. Sketching NLP: a case study of exploring the right things to design with language intelligence. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–12.

63. Qian Yang, Alex Scuito, John Zimmerman, Jodi Forlizzi, and Aaron Steinfeld. 2018. Investigating How Experienced UX Designers Effectively Work with Machine Learning. In *Proceedings of the 2018 DIS Designing Interactive Systems Conference*, 585–596.

64. Qian Yang, Aaron Steinfeld, Carolyn Rosé, and John Zimmerman. 2020. Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–13.

65. Anon Ymous, Katta Spiel, Os Keyes, Rua M. Williams, Judith Good, Eva Hornecker, and Cynthia L. Bennett. 2020. “I am just terrified of my future”—Epistemic Violence in Disability Related Technology Research. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–16.

66. Yuhang Zhao, Shaomei Wu, Lindsay Reynolds, and Shiri Azenkot. 2018. A Face Recognition Application for People with Visual Impairments: Understanding Use Beyond the Lab. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 215.

67. John Zimmerman, Jodi Forlizzi, and Shelley Evenson. 2007. Research through design as a method for interaction design research in HCI. In *Proceedings of the 2007 CHI Conference on Human Factors in Computing Systems*, 493–502.

68. Responsible AI Principals from Microsoft. Retrieved from https://www.microsoft.com/en-us/ai/responsible-ai