THE UNIVERSITY OF CHICAGO

A HUMAN-CENTERED APPROACH TO AI FOR SCIENCE AND HEALTHCARE

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF COMPUTER SCIENCE

BY
KELLY BARBARA WAGMAN

CHICAGO, ILLINOIS
JUNE 2025

Abstract

Artificial intelligence (AI) is increasingly integrating into everyday life, with significant potential to transform healthcare and scientific research. These fields offer unique opportunities for societal benefit—including improved health outcomes, climate adaptation, and accelerated scientific discovery—but successful integration of AI requires a sociotechnical approach that takes into account human norms and practices. This thesis explores how AI can be designed and deployed in human-centered and inclusive ways, particularly for historically marginalized communities. Through three case studies, I investigate AI's role in real-world settings: (1) examining how Black older adults socially relate to an AI voice assistant for home exercise, (2) analyzing the adoption of the first generative AI chatbot in a U.S. national laboratory from an organizational perspective, and (3) developing a participatory AI system for climate adaptation that integrates community-driven data. This work contributes to understanding how to design effective and inclusive AI systems in science and healthcare settings.

In memory of my dad Charlie

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

PREFACE

I want to begin by explaining my positionality in relation to the research in this thesis. I started my career as a software engineer at a large tech company, where *scale* and *productivity* were core drivers of the products developed. I realized much of the tech industry was built around these concepts, since they can lead to successful businesses. I wanted to know what kinds of technologies were not being built due to this focus. I was not satisfied, however, with work critiquing the tech industry by pointing out flaws without proposing other ways of building. I felt that the space of possible artifacts and sociotechnical configurations was larger than what was presented with the current set of available products. Also, as a white woman, I wondered about the lack of gender and racial diversity in the US tech sector and how this played a role in limiting the kinds of products and businesses that got built.

To answer these questions, during my MS at MIT in Comparative Media Studies, I dug into the ideology and logic underlying much of the tech industry, explored alternatives, and developed a strong theoretical basis for building technology grounded in feminist ideals [200]. I called this theoretical framework *relational engineering*, defined as *"a technology design ethos that prioritizes the development of caring relations in a sociotechnical system throughout the lifetime of the artifact/system... This mindset encourages designers to craft relations as opposed to objects."* In my MS thesis, I argued that approaching technology design this way would yield different outcomes than those traditionally found in the tech industry. While I had several examples, I did not have time to actually try to run a full project guided by relational engineering.

For me, a PhD in Computer Science was a chance to test relational engineering as a framework from first principles. Shielded from the need to produce profitable products, was it possible to create technology by centering the development of caring relations? It was important for me to understand this in the context of messy, real-world settings since I wanted to know if building technology this way could work outside an academic context.

AI has been a particularly interesting technology to me for a long time because of its tendency to enmesh so deeply with human social structures. Humans constantly create social structures that influence how we act; some structures are temporary while others are lasting, such as those formalized into institutions. AI that can mimic human communication (e.g. speech) insert themselves into these structures by adopting an identity that other social actors then perceive and relate to. AI infrastructure also intertwines itself by collecting the traces made by humans and machines (data), reconfiguring it, and codifying some configurations and not others. In my MS thesis, I also developed the theoretical groundwork for building equitable and inclusive AI systems by defining a *social machine* model [200, 201]. Building on relational engineering and the social machine model, I wanted to understand if it would be possible to design AI systems that benefit human social structures through caring human-machine relations.

At the beginning of the PhD, with my theoretical framework in hand, I was faced with the task of figuring out where to build and who to build with. After exploring a number of directions, I settled on focusing on applications in science and healthcare. Of all the ways AI is changing sectors, I felt (and still feel) that some of the most societally important breakthroughs will come from applying AI to healthcare and scientific research.

After four years of applying a relational engineering perspective, I have found that while it is hard and time consuming it is also possible and rewarding. Moreover, I have been able to (viscerally) understand how it presents a strategy not only for building technology, but for collaborative worldbuilding. This thesis presents the academic findings from the research. But I also want to acknowledge the power of knowing in my gut I can create new technology (and micro worlds) starting from care and relationship building.

ACKNOWLEDGMENTS

I am grateful to the large village that has supported me throughout the PhD process. To my advisor Marshini Chetty, thank you for always listening to my half-formed ideas, for your endless encouragement in the face of challenging circumstances, and for helping me find the path forward when I was lost. To the other students in the AIR lab, including Jake Chanenson, Lan Gao, and Brennan Schaffner, thank you for creating an inclusive and supportive lab culture and for listening to many practice talks.

All of the projects in this thesis included collaborations both within and outside the UChicago computer science department, and I am deeply grateful to my partners for their time and insight. To Nedra Sims Fears, the Greater Chatham Initiative, and the residents of Chatham, thank you for generously inviting me into your community and spending time designing *Water On My Block* with me. To Thomas Chang, Bonnie Ko, Kanchan Naik, John Rugemalila, and Madison Vanderbilt thank you for the many hours you've spent researching, designing, and bringing *Water On My Block* to life, I can't wait to see what you go on to build. To Scott Collis, Robyn Wheeler Grange, Jorja Porter, and the CROCUS team, thank you for believing in my vision for how human-computer interaction could contribute to climate science and helping me overcome hurdles to achieve that vision. To Megan Huisingh-Scheetz and the EngAGE clinical trial team, thank you for allowing me to conduct qualitative research with clinical trial participants and helping with the complex logistics needed to make that possible. To Matthew Dearing, thank you for your trust and support in studying Argo. To Arthur Borem, thank you for being an incredible paper writing partner through challenging times. To Tamara Clegg and Jessica Vitak, thank you for helping me become a better writer and researcher. To Nancy Baym, thank you for your mentorship over many years and for nurturing my early interest in AI and the social world. To the mentors I have been lucky to have at UChicago, including Ian Foster, Alex Kale, Mina Lee, Pedro Lopes, and Sarah Sebo, thank you for your support and feedback throughout the PhD.

Graduate school also requires significant personal support. In the middle of the PhD, I lost two of the most important people in my life: my dad Charlie and my grandfather Bill. I attribute my interest in solving practical, real-world problems to my dad, a life-long small business entrepreneur who was always one step ahead in applying new technology in practical ways. My grandfather was always interested in how my research fit into global macroeconomic trends, which pushed me to be able to explain the relevance and impact of my work. I am grateful to the other members of my family who have always cheered me on: my mom Reid, Meagan, JJ, and Bob, as well as my extended family including Kathy, Nicky, Ed and Beth. A huge thank you is also due to my husband Mike, who has always believed I can do hard things, and given me unconditional support to go do them.

Lastly, thank you to the baristas at *Maple Leaf Coffee House*, where much of this thesis was written.

CHAPTER 1

INTRODUCTION

Artificial intelligence (AI)—including technologies such as machine learning (ML) and large language models (LLMs)—is "reaching out" of research labs [13] and becoming increasingly integrated into the lives of everyday people. While there are many possible applications of AI in the real world, healthcare and scientific research present particularly appealing opportunities because improving healthcare and advancing the rate of scientific discovery has the potential to provide direct benefits to society such as mitigation of disease, adaptation to climate change, and acceleration towards the use of renewable sources of energy. While advancements in these areas require technical innovation and the deployment of new technologies, these are situated within social contexts that must support and adopt the technical changes in order for them to be effective. Therefore, as we consider the possible transformative future of AI in science and healthcare, it is critical to understand and design for this change through a sociotechnical and human-centered lens. Moreover, to ensure these benefits reach everyone in society, it is important to understand how people from historically marginalized backgrounds can be included in the transformation. This is because historically marginalized groups have often been ignored when designing new technology—or worse, harmed by it—and there is less research on perspectives from these groups (see e.g.[72, 18]).

In this thesis, I document three projects where I investigate the integration of novel AI systems into healthcare and science applications using methods from the field of human-computer interaction (HCI), while paying particular attention to the inclusion of historically marginalized voices in these systems. Taken together, these projects provide a series of case studies for integrating AI into real-world healthcare and science scenarios using a sociotechnical, human-centered perspective and with an emphasis on building inclusive systems. In the first project, I study the deployment of an AI voice assistant to support Black older adults in exercising at home. In this study, I look, in particular, at how participants related socially to the AI personality and argue that this social relationship informs participant willingness to listen to the health advice from the AI. In the second project, I study the initial deployment of generative AI in a U.S. national lab and outline both use cases and risks. Here, I take an organizational perspective and study adoption by both the Science and Operations sides of the lab since these are both important to the day-to-day success of scientific research. In the third project, I design and build a participatory AI system for climate adaptation that simultaneously collects community-based data for inclusion in climate ML models and provides an interface for community members to directly benefit from this data. My thesis both contributes to academic research and has real-world impact.

In the remaining sections of the Introduction, I will provide a background on sociotechnical and human-centered design, why these methods apply to science and healthcare contexts, and give a chapter outline.

1.1 Background on a Sociotechnical and Human-Centered Approach to AI System Design

This thesis uses a sociotechnical perspective on human-centered technology design in order to make the design of AI systems more inclusive and beneficial. In this section I will define what a sociotechnical system is and how it relates to human-centered design and similar design theories, and how these apply to the design of AI systems.

1.1.1 Sociotechnical Systems Thinking in HCI

A sociotechnical system is an assemblage of the social and technical. Technical artifacts involved in a system might include objects such as an iPhone, as well as computer code, AI models, rare earth minerals, and more; social aspects of the system include cultural norms and values, power dynamics, and other forms of social relations [52]. An important part of a sociotechnical systems perspective is understanding that the technical aspects of a system influence the social aspects, and vice versa. That is, the social and technical are mutually shaping [201]. As an example, the introduction of the mobile phone was influenced by existing ways that peopled interacted through voice, and the technological development of texting in turn shaped interpersonal communication norms and practices [16]. Introducing a new technology into any sociotechnical system will send ripple effects throughout.

Sociotechnical thinking has been woven into HCI, computer-supported cooperative work (CSCW), and related disciplines. One way that sociotechnical ideas entered HCI was through participatory design, particularly in Scandinavian approaches that emphasized democratic involvement in technology development [70]. At the same time, CSCW researchers recognized the need to address both technical and social aspects of systems in the workplace (later branching into other domains). Scholars like Lucy Suchman challenged traditional cognitive models in HCI by emphasizing the contextual and social nature of human-computer interactions [187]. In the early 2000s, researchers began explicitly framing HCI as a sociotechnical problem. For example, Mark Ackerman highlighted the gap between social needs and technical feasibility [4]. More recently, sociotechnical thinking in HCI has informed work on critical computing, AI ethics, and justice-oriented design. For example, Paul Dourish has explored embodied interaction [65] and Sasha Costanza-Chock has studied how systems reinforce and challenge power structures [49]. This thesis takes sociotechnical thinking beyond reflecting on existing systems towards seeking to apply this way of understanding technology in the design of new systems.

1.1.2 Human-Centered Design and Related Frameworks

The field of HCI has been built around the idea that technology should address real human needs, and not simply proliferate cool gadgets. In the 1980s, Don Norman coined the term "human-centered design" as a way of focusing designers' attention on the needs of the end user rather than becoming overly fixated on technological capabilities. This concept was further popularized in his book *The Design of Everyday Things* [128], where he emphasized the importance of usability, cognitive affordances, and intuitive design. A human-centered design framework begins by asking what people need or want and then shaping technology accordingly, rather than forcing users to adapt to poorly designed systems. The process typically involves iterative design, user research, and testing to ensure that technology aligns with human behavior.

More recently, scholars have pointed to the fact that Don Norman's original conception of human-centered design focused on usability by a singular end user, and did not factor in values beyond usability such as equity and inclusion and as well as broader sociotechnical impacts. Numerous design frameworks have been developed that offer strategies for emphasizing inclusion, equitable power dynamics, and reimagining the role of designed objects. For example, participatory design integrates stakeholders into the design process and has evolved into human-centered and inclusive design approaches [49]. Value-sensitive design (VSD) highlights how designers' biases can be embedded in the technology they make and offers strategies and tools for designers to recognize this and reduce bias [74, 75]. Critical and speculative design are methods for challenging societal assumptions through conceptual artifacts, for example via designing objects that might exist in a science fiction future as a way of imaging and questioning that future [67] or using objects for political critique and democratic engagement [64]. Finally, feminist and justice-oriented design approaches, such as critical fabulations [158], feminist HCI [14, 15], data feminism [69], and design justice [49], work to reshape narratives, expose inequalities, and advocate for restorative design practices.

Building on work on sociotechnical systems in Science and Technology Studies (STS) and inclusive design frameworks such as feminist HCI, I have argued in prior work that it is critical to approach design and engineering problems by centering the relationships built throughout the process as opposed to simply the artifact [201, 200]. In other words, starting from the perspective that the relations are the primary design focus. These relations encompass not just the relationship between the end user and the final artifact, but the relations developed along the entire life cycle of the artifact from conception through deployment. This differs from most of engineering practices that exclusively center the final object and its use by a primary end-user.

1.1.3 Applications to AI System Design

Building on the idea of "human-centered design," scholars have argued for "human-centered AI" (see e.g. [175, 9]) that reinforces the idea that AI must be designed not just for efficiency, but with regard to its users and social contexts. Human-centered AI has evolved from a desire to produce systems that do not cause harm and that add value to human lives; examples include diverse issues such as mitigating bias in AI, AI-powered disinformation campaigns, realizing business value from generative AI, and existential threats posed by AI [214]. A review article on human-centered AI literature found that papers on this topic could be grouped in four main categories: 1) explainable and interpretable AI, 2) human-centered approaches to design and evaluate AI, 3) humans teaming with AI, and 4) ethical AI [41].

Sociotechnical studies of AI emphasize AI as a product of political, economic, and social forces. Scholars in this field have critically examined how AI systems shape and are shaped by broader societal structures, exposing issues related to power, labor, surveillance, and environmental impact (e.g. [86, 51]). For example, Kate Crawford's *Atlas of AI* [51] maps how AI depends on extractive processes such as mining rare earth minerals, exploiting low-wage labor, and harvesting massive amounts of data to reveal how AI's development is tied to histories of colonialism, capitalism, and inequality. She critiques the idea of AI as an autonomous system and instead frames it as a sociotechnical network driven by human and institutional interests. Other sociotechnical critiques of AI explore themes such as racial inequities [18, 34, 127] and socioeconomic inequities [72, 86] that reinforce existing patterns of marginalization.

This range of scholarship on human-centered and sociotechnical AI begs the question, what is AI? While there is no singular definition and it is context dependent, for the purposes of this thesis I will use the following two-part definition. First, AI is a system that instantiates a function. In other words, an AI system has an input, processes that input using an algorithm, and has an output. Included under this umbrella is rule-based AI, machine learning, and large language models. Second, AI is a system designed to interact socially with humans. Examples of social interaction might include human-AI interfaces such as chatbots as well as sociotechnical interactions such as human data collection mechanisms for AI systems or AI-assisted decision-making systems. Due to the ambiguity of the term AI, in prior work I proposed using the term "social machine" instead and argued that considerations of equity and inclusion should be central to designing social machines: a "social machine" is *"an object that is designed to construct and engage in social relations with humans, and that has been crafted with careful attention to issues of agency, equitability, inclusion, and mutuality"* [201]. In this thesis, I build on this prior work and continue to advocate for designing social machines that are equitable and human-centered. I use the term AI instead of social machine since it is more widely used in current academic literature.

1.2 Human-Centered AI in Science and Healthcare

AI, particularly machine learning (ML), has become increasingly important in science and healthcare. Researchers are using AI to find patterns in genome datasets, understand interactions between subatomic particles, and discover new drugs (e.g. [182, 194, 8]). However, as AI systems become more deeply embedded in scientific and healthcare workflows, they are no longer just analytical tools but can be part of decision-making and discovery processes. This shift makes human-centered AI essential for ensuring these systems are effective, trustworthy, and aligned with human needs.

As discussed in the prior section, AI operates within complex sociotechnical systems and this is also true in science and healthcare contexts where social norms and practices shape outcomes. In healthcare, for example, AI can complement human clinical judgment in diagnostics and treatment (e.g. [17, 38, 39, 73, 97, 135, 215]). In scientific research, AI models might influence how knowledge is produced or automate scientific workflows (e.g. [182, 194, 8, 62, 115, 150, 164]). In addition, AI systems increasingly rely on sensitive human data, such as patient records, biomedical images, and clinical trial results. This raises concerns about privacy, bias, and fairness, particularly for groups who have been historically underrepresented in medical research and may experience disparities in AI models and their applications (e.g. [34, 43]). Given the vast domain of science and healthcare, more work is needed in order to understand how to apply AI in these areas as effectively as possible in ways that compliment and benefit humans.

To expand the body of research on sociotechnical approaches to AI design in science and healthcare, I focus on three case studies in this thesis. These case studies can be segmented by three sociotechnical contexts: the home, the organization, and the community. My first project is situated in the social context of the home: I examine how Black older adults use a voice assistant for exercise at home. The home is a unique social context because people typically consider it private space where they can be themselves and act on the norms and cultures they are most comfortable with. When the telephone was first introduced to the home, people became concerned because it created a direct line for the outside/public world into the private sphere of the home [119]. In this project, I pay particular attention to how people relate to a conversational AI agent in their home and privacy concerns they may have. My second project is situated in an organization focused on scientific knowledge work. Organizations are social entities that have their own cultures, norms, and practices. Organizations are sociotechnical because they are assemblages of these social relations along with numerous technical artifacts that influence and mediate social relations. Understanding how generative AI is impacting scientific knowledge work from an organizational perspective can influence how this type of AI is designed in ways that make the scientific process more efficient. Lastly I focus on the social context of a community. There are many kinds of communities; in this work I look at a co-located community representing a neighborhood on Chicago's South Side. In this case study, climate scientists are building AI models to simulate weather/climate in the neighborhood and are interested in community-sourced data. This project deals with the complexity of building AI systems with communities and what needs to be done to ensure the models, and data collected, ultimately benefit the people in the community. While these case studies focus on specific groups at the home, organization, and community level and do not provide representative large-scale samples, they are able to deeply interrogate what it means to build science and healthcare AI in social contexts and the insights and contributions of each project are broadly applicable.

1.3 Chapter Overview

In Chapters 2—4 I report on three case studies focused on a human-centered approach to AI for science and healthcare. These three projects illuminate ways human-AI interaction can be understood and improved in healthcare and science settings. Namely, the first two projects highlight ways in which human-AI collaboration is nuanced and getting these relationships right in the design of systems will impact their efficacy and usefulness to ultimate goals in health or science. The third project shows that designing beneficial participatory AI systems is challenging and time-consuming, but can lead to meaningful multi-stakeholder benefits if well-executed. In Chapter 5 I provide directions for future work and summarize the contributions of this thesis.

In Chapter 2, I look at human-AI interaction in healthcare with a focus on Black older adults using a voice AI at home for exercise. I led a team that studied an existing conversational AI program designed by medical researchers to help older adults exercise at home using an Amazon Echo Show (Alexa) device. The medical researchers ran a clinical trial where one participant group has access to the exercise program (called EngAGE) on the voice assistant device while the other group has the same exercises in a paper packet. The trial specifically focused on Black older adults in lower-income neighborhoods around Chicago. Our team conducted two qualitative studies (N=34 total), one before and one after the clinical trial, to understand 1) how participants socially related to Alexa and how this informed their use of the device for health, and 2) what factors motivated participants to exercise and which of these factors derived from the voice assistant device versus broader sociotechnical factors such as support from friends and family. Overall, this research showed how a conversational voice assistant can augment health motivation for some (although not for others who relate to the virtual agent as a stranger or intruder), but human social support and other sociotechnical factors are still critical. This is important as health advice and personalized healthcare may increasingly be communicated through AI agents.

In Chapter 3, I report on a study of the deployment of a generative AI chatbot in a professional science research organization [202]. For this project, I led a team that used an organizational perspective to study the rollout of the first generative AI chatbot to Argonne National Lab in order to understand both use cases and risks. We conducted a survey (N=66) and interviews (N=22) with employees across Science and Operations roles at Argonne to understand how generative AI is shaping knowledge work in this context. Given the confidential nature of some national lab research, we pay particular attention to concerns surrounding privacy, security, and ethics in this setting. This research showed that professionals in a science organization are increasingly using generative AI, that these use patterns fall into either a copilot or workflow agent modality, and that there are a number of features specific to science organizations in particular that are at risk with generative AI such as the need to maintain public trust. These findings are important because generative AI is rapidly being integrated into workplaces, including science organizations, and this research can contribute to shaping how the technology will evolve along with the guardrails that are needed.

In Chapter 4, I led a team that collaborated with professional climate scientists and a community non-profit in Chicago to use a human-centered design process to construct a participatory AI system. The project built on an existing initiative at Argonne National Lab to develop hyper-local climate AI models designed specifically for Chicago neighborhoods, which involved placing weather station nodes in these neighborhoods. We wanted to investigate how Chicago communities, particularly those that have been historically overlooked, could participate in, and benefit from, these climate models. To do this, we embarked on a complex, multi-year process to build relationships with, and align, key stakeholders in climate science and community organization. Climate concerns in Chatham are also public health issues because, for example, flooded basements lead to mold and poor air quality exacerbates conditions like asthma. Through many discussions, in-person events, interviews (N=15), and focus groups (N=62) throughout the course of over two years led to the development of a flood reporting app for the Chatham community. The app's primary goal was to benefit the community by allowing residents to report street and basement flooding; the community non-profit we collaborated with owns and controls the data collected. In addition, an anonymized version of the data collected can improve the climate AI models being developed by scientists. This project contributes insights on the limits and benefits of developing participatory AI in a real-world science setting.

In Chapter 5, I will review the contributions of this thesis and provide directions for future research. In particular, I will discuss future work in designing culturally-sensitive interpersonal relations with AI, developing copilots for organizational information retrieval, and fostering collaborations with community data partners.

CHAPTER 2

EXAMINING BLACK OLDER ADULTS' PERCEPTIONS AND USE OF DIGITAL VOICE ASSISTANTS FOR EXERCISE AT HOME

2.1 Introduction

In this chapter[1] I discuss a study I led on using voice assistants for older adults to exercise at home. Voice assistant (VA)[2] technologies such as the Amazon Echo (Alexa) and Google Home are promising tools for supporting older adults with health-related activities [170, 104, 24, 44, 132, 30, 103]. Preventative health measures,[3] such as physical activity, are critical for older adults [36, 35]: a significant body of medical literature has found that exercise in particular improves many facets of older adults' health including frailty, sleep, blood pressure, reduced risk of falling, heart disease, some kinds of cancer, dementia, and depression and anxiety [92, 133]. In addition, minoritized communities[4] often have worse health outcomes and face obstacles to receiving care [197, 10]. Voice assistant devices with preventative health applications offer one possible direction for providing healthcare in the home, but it is unclear how effective this strategy might be at motivating preventative health behaviors such as exercise for minoritized groups.

Due to the demonstrated need for providing minoritized communities with accessible

1. Sections of this chapter are under review as a paper. Co-authors on the paper include Arthur Borem, Lan Gao, and Marshini Chetty. Special thanks to Megan Huisingh-Scheetz and the entire EngAGE team for their support.

2. In this chapter we use the phrase *voice assistant* because that has become a widely used term. We do not imply that these devices fulfill the role of an "assistant" and later describe the range of relations participants have with the device. We more closely mean *digital voice agent*.

3. In this chapter when we use the term *preventative health* we refer to physical health as opposed to mental health.

4. Following prior work in HCI (e.g., [139]), we use José Muñoz's term *minoritized* [124] to highlight a power imbalance between a community and a social majority that may not necessarily stem from underrepresentation.

preventative care, in this chapter we seek to understand to what extent voice assistants can be effective in encouraging Black, largely lower-income older adults in a large US city to exercise at home. Prior work has found that while older adults tend to have a favorable opinion of voice interfaces, including in a healthcare context, over time they often stop using VAs due to issues such as reliability concerns, lack of trust, and usability challenges [219, 44, 54, 184]. Within health applications for older adults, many VA studies have focused on health information seeking using a VA device [148, 29, 88, 122, 20]. Some studies look specifically at Black older adults, finding concerns such as VAs not understanding African-American Vernacular English [88, 31], while others have looked at lower-income communities [125, 89, 181]. We expand on this work by focusing on preventative health maintenance using VAs, understanding perspectives from Black older adults in a lower-income community, and conducting both in-person focus groups and a comparative study of long-term (six month) use of a VA compared to a similar non-technological intervention.

To conduct our study, we collaborated with geriatric medical researchers who are conducting a clinical trial[5] testing the efficacy of an exercise app (called EngAGE) on a VA device. The clinical trial specifically recruited Black older adults in a large US city, many of whom live in lower socioeconomic areas of the city. EngAGE was designed by the medical research team for the Amazon Echo Show device, and uses voice to explain each exercise and also shows a static image of what the exercise looks like on the screen. Users can provide verbal feedback on if the exercises are too easy or too hard and the program will adjust accordingly. The exercises do not require any specialized equipment. In the clinical trial, participants are randomized into two groups. The first group has access to EngAGE on the VA. The second group has access to a paper packet covering the same exercises using identical language and photos. Participants are in the clinical trial for a six month period, providing extended access to the VA. See Figure 2.1 for the study timeline.

5. The clinical trial is ongoing and therefore we cannot publish any trial results.

I led a team in conducting two qualitative studies pre- and post-clinical trial. During these studies we focused on two main research questions:

RQ1: How do Black older adults *perceive* voice assistants and how do these perceptions influence voice assistant effectiveness for promoting preventative health?

RQ2: What *motivating factors* influence Black older adults' decision to exercise, and to what extent is a voice assistant exercise app motivating relative to paper exercises?

In the first study, conducted prior to the start of the clinical trial, we held in-person focus groups with Black older adults (N=15) that included live interactions with Alexa via an Amazon Echo Show device. In these focus groups, we wanted to understand participant perspectives on interacting with the device and to what extent these perspectives influenced envisioned use, especially for preventative health applications. In the second study, we conducted interviews with clinical trial participants[6] (N=20) immediately after they finished their six month study period, where half of the participants had access to the VA exercise program and half had a paper packet of the same exercises. In these interviews, we wanted to understand what factors influenced participants to exercise and, in particular, what differences there were between the VA and paper packet regimes.

In Study 1, we found participants held a wide variety of differing perceptions about using Alexa, which in turn impacted their willingness to use the device for health purposes. To capture this heterogeneity and situate the findings in the cultural context of the study population, we organize participant perspectives into five interpersonal relationship types: tool, amusement, friend, stranger, and intruder. Understanding these interpersonal relations is critical in a preventative health setting since we found older adults with negative perceptions were uninterested in using the device for health purposes.

In Study 2, we compared motivation to exercise between the group of participants using the VA and the group using the non-technological exercise instructions (a paper packet).

6. One participant was in both studies.

We describe each of these motivating factors and, to organize them and tease out the effect of VA use, we sort them into four interaction levels: individual (e.g., intrinsic motivation), human-artifact interactions (e.g., Alexa's personality), social support interactions (e.g., care partner support), and research study interactions (e.g., tech support offered by clinical trial staff). Our key finding is that a significant amount of motivation to exercise comes from human social support, regardless of the use of a VA or paper exercise packet. Combining the results of Study 1 and Study 2, we found that the VA could provide additional motivation to exercise but only if the participant had a positive interpersonal relation with it. This chapter makes the following contributions to the HCI and CSCW community:

- We extend work on Black and lower-income older adults' use of voice assistants (e.g. [88, 31, 125, 89, 181]) as well as work on the culturally-informed interpersonal dynamics between users of all ages and VAs (e.g. [147, 80, 163, 216, 154]), showing that people in our study population have heterogeneous social perceptions of the VA that include tool, amusement, friend, stranger, and intruder and these relational types impact participant willingness to use the device for health purposes.

- We extend work on preventative health maintenance technologies (e.g., [46, 212]) and older adult use of VAs for tasks like health information seeking (e.g. [148, 29, 88, 122, 20]) to focus on to what extent VAs can motivate participants to exercise, breaking down motivation into 11 factors across four interaction levels.

- We conduct our studies in collaboration with a clinical trial team in order to both run focus groups in-person with participants (Study 1) and to work with participants who had been using a VA or a paper exercise routine at home for six months (Study 2), providing a long-term comparative study in HCI/CSCW of voice assistant use by older adults in a healthcare setting.

2.2 Related Work

In this section, we review prior research on overall adoption of voice assistants by older adults, voice assistants in healthcare for older adults, and technology-driven preventative health maintenance, respectively.

2.2.1 Adoption of Voice Assistants by Older Adults

Researchers in CSCW, HCI, and related fields have been interested in the adoption of voice-enabled technologies for older adults since these technologies do not rely on graphical or physical user interfaces which may be challenging for people with low vision and decreased dexterity (e.g., [99, 196, 184]). Much of this research is focused on usability by older adults [99, 184, 12, 23]. Overall, studies have found that while older adults have favorable initial attitudes towards voice interfaces, there are a number of barriers to use, especially for stand-alone voice assistant devices such as the Amazon Echo. For example, one study of older adults (predominately higher-income with Master's degrees) who used voice assistants for over a year found that participants abandoned the device over time due to finding few beneficial use cases, concerns that the VA might limit independence, and difficulty navigating use in shared spaces [195]. Similarly, a three-week field trial for older adults with the Amazon Echo Dot found a disparity between the perceived usefulness of the device for tasks such as reminders and actual use due to fears of unreliability [148]. Older adults also have concerns about being patronized or manipulated by VAs [94, 173]. Several studies have outlined breakdowns[7] in older adult–virtual agent systems due to communication (VA requests require a particular type of structured language that can be challenging for older adults) and maintenance challenges that can lead to non-use [88, 148, 217, 107]. Other scholars have investigated older adults' privacy concerns with respect to voice assistants and how

7. We use the term *breakdown* from the field of Science and Technology Studies to denote an aspect of a sociotechnical system that causes friction or does not work as intended; breakdowns often require maintenance or repair work (see e.g., [98]).

this impacts adoption [22, 60, 106, 94, 173, 40].

While there have been studies of older adults and VAs, these studies have largely not focused on minoritized older adults specifically. A review paper on digital technology and older adults argues that most research has treated older adults as a homogeneous group while ignoring their different age ranges and cultural backgrounds [143], thus it is critical to examine a variety of older adult perspectives. Three additional review papers looking at older adult use of voice technologies [99, 138, 171] either never mention or make only a single reference to race, ethnicity, or cultural background. Further, the review paper by Sezgin et al. [171] specifically states that "*VA developers will need to consider the contextual setting for the behavioral health interventions*" but that this had not been done.

Research on Black older adults' use of voice assistants has pointed to the need for voice assistants to understand and speak African-American Vernacular English (AAVE). In one paper, Harrington et al. [88] interviewed Black older adults residing in lower-income environments about using the Google Home voice assistant to search for health information. They found that since the voice assistant does not always understand AAVE, older adults in this group needed to "code switch" and alter their language in order to communicate with the device. Similarly, Bosco et al. [26] studied Black older adults searching for health information on dementia and found that voice assistants "*have a tone of voice that resembles African American and Black voices.*" Additionally, Brewer et al. [31] studied how Black older adults' conceptualized "AI fairness" in voice technology, unpacking how their participants envisioned AI-driven voice assistants that could be more inclusive in terms of speech patterns and content.

Additional research has looked at the use of voice assistants by people from lower-income communities. Studies have raised concerns over both the cost of the device and the Internet connection needed to run it [125, 89]. So et al. [181] conducted a study on health-focused voice assistants with older adults from a range of racial/ethnic backgrounds in a lower-income neighborhood and found that participants wanted a "non-judgmental" device that could securely discuss stigmatized topics including drug use, isolation, and dementia. Health data confidentiality was also a significant concern for participants [125, 181].

Our paper contributes to the literature on the adoption of voice assistants by Black older adults in lower-income communities. In particular, we expand research on topics like speech patterns to include other social and cultural factors—such as interpersonal relations with the VA—that are important for VA adoption by this population.

2.2.2 Voice Assistants in Healthcare for Older Adults

In addition to general adoption of voice assistant technology by older adults, prior research has studied how voice assistants might aid older adults in a healthcare context [170, 104, 24, 44, 132, 30, 103]. One body of research investigates how older adults use voice assistants and conversational agents to search for health-related information from home [148, 29, 88, 122] as well as from a hospital [20]. For example, research has looked at the extent to which older adults trust VAs and the health information they receive from the VA [87, 104, 54]. Other research studies the use of voice assistants by people, including older adults, for accessibility purposes [218, 146] and for tasks such as medication management [120]. Similar to findings on VA adoption described earlier (Section 2.2.1), researchers found older adults commonly struggle with usability limitations in VA healthcare systems [219, 44]. To address these limitations and barriers to use, some research has used co-design methods with older adults and caregivers for creating a more inclusive, accessible, and ethical voice assistant for health-related activity [95, 210, 112, 181].

Research on minoritized groups in a healthcare context is important since these groups face negative health outcomes and barriers to care [131, 153, 10, 87]. For instance, research suggests VAs have the potential to reach people who may have difficulty getting to an in-person healthcare location [122]. A review paper [113] on utilizing conversational agents in healthcare highlighted that VA healthcare applications could be subject to bias and reinforce stereotypes that may negatively impact particular groups.

Research on the use of VAs in healthcare for older adults could be expanded to include more work on *preventative health* specifically. Preventative health measures are activities that reduce the likelihood of future disease, such as physical activity and a healthy diet. In one paper, Hu et al. conducted co-design sessions with older adults about designing a VA application for well-being and used an exercise-focused probe [95]. Bickmore et al. tested an exercise-focused conversational agent with eight older adults over a two month period, with participants having a generally positive view of the agent [19]. A review paper by Sezgin et al. [171] reported that there was evidence VAs could decrease pediatric body mass index (BMI) z-scores, but there is not an equivalent study for older adults.

We contribute to the conversation on the use of VAs by older adults in a healthcare context by focusing specifically on exercise as a form of preventative health in a population of Black older adults in a lower-income urban environment. Due to the fact that we have collaborated with a clinical trial that has randomized participants into a VA exercise group and a paper-based exercise group we are able to identify novel insights pertaining to exercise motivation and use by comparing participant experiences between trial groups.

2.2.3 Technology-Driven Preventative Health Maintenance

Outside the scope of voice assistants and older adults, researchers have long conducted studies on how technology more generally might promote awareness of, and action toward, preventative health maintenance. One body of research investigates the potential of current digital technology to drive health behavior change. For example, Conroy et al. [46] analyzed 167 top-ranked mobile phone applications focused on physical activity in order to identify which behavior change techniques each app used (e.g. providing instructions on exercise, modeling how to perform an exercise, providing feedback, goal-setting for physical activities, and planning social support), finding that most apps did not use many of these techniques. Some research has looked at preventative health maintenance for specific populations such as people with chronic diseases [91, 178, 168] and people in lower socioeconomic groups [209].

CSCW and HCI researchers have evaluated the user experience of digital tools for preventative health maintenance. For example, through a 28-day field study, Xu et al. [212] investigated how users planned exercise through physical activity mobile applications, as well as how they succeeded or failed to integrate their plans into daily life. Some work has implemented visualizations to encourage users' behavior by helping them reflect on their physical activities effectively and intuitively [206, 207]. Other work has utilized technology to facilitate intrinsic and social motivation in preventative health, for example by promoting reflection [213] and peer or stranger social support [5, 6, 47, 130].

We build on initial work to understand technology-driven tools to assist older adults' physical exercise (e.g., mobility) [45, 100, 108] and prior work on preventative health-focused mobile phone apps. In addition, we add a longer-term preventative health study to this literature to investigate how voice assistants could aid minoritized older adults.

2.3 Methods

All procedures were approved by our Institutional Review Board (IRB).

2.3.1 Research Context

In this study we (HCI / Computer Science researchers) collaborated with medical researchers specializing in geriatric care who were running a clinical trial to test the efficacy of an exercise app for the Amazon Echo Show device for older adults. The medical researchers developed and successfully piloted the app, called EngAGE, prior to the start of our collaboration.[8]

8. A full discussion of EngAGE pilot results is beyond the scope of our study.

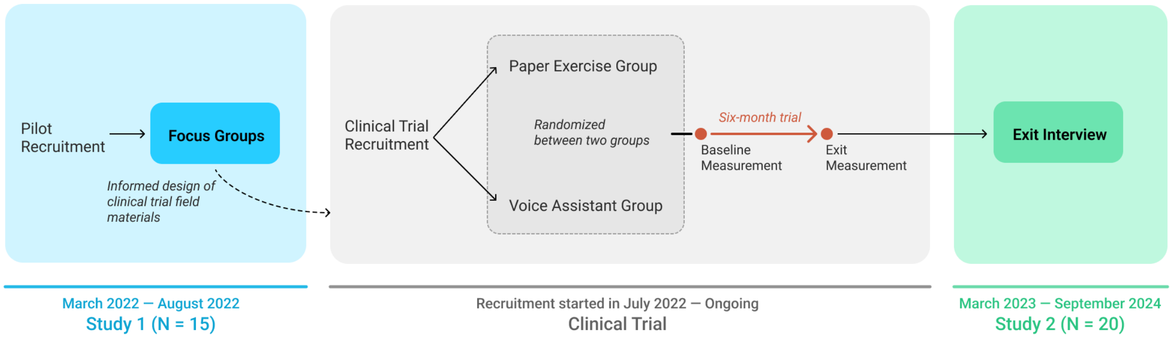


Figure 2.1: Timeline of our Study 1 and Study 2 and where they fit within the clinical trial period (this chapter does not report clinical trial results, as the multi-year trial is ongoing).

EngAGE helps older adults with mobility exercises. It walks users through a series of exercises that do not require specialized equipment, such as toe stands or bicep curls with soup cans, using voice instruction and static images. Users can provide verbal feedback to make the exercise routine harder or easier. App users can also provide care partners with access to their exercise history and care partners can leave motivating messages.

The medical researchers are conducting a clinical trial to test EngAGE (see Figure 2.1). The clinical trial targeted participants who lived in the Chicago area, identified as Black, were over 65 years old, had multiple health issues, and were largely homebound (i.e., cannot leave the house because of age or health issues). Older adult participants were recruited along with a designated care partner, who could be a family member or friend. The care partner was instructed to help the older adult remember to do the exercise routine. The clinical trial had two research groups and each participant was randomly assigned to one of these groups. In the first group, participants were given an Amazon Echo Show set-up with the EngAGE. In the second group, participants were given a packet of paper with the same exercises (exactly the same images and text as in the app version). Participants in both groups received printed instructional materials and were given a phone number for a clinical trial full-time staff member who could answer questions and provide technical support as needed. The EngAGE exercises differed from the paper exercises in that the instructions were spoken through the Echo Show speaker and that the device had a built-in timer for each exercise that was set based on oral feedback provided by participants indicating the difficulty of the exercises (for example, if exercises were "too difficult" the program would reduce the exercise time period). Participants were asked to do the exercises over a period of six months. Health metrics were measured by the clinical trial team at baseline before the trial and at the end of the six month trial period.[9]

We conducted our Study 1 before the clinical trial started and Study 2 with participants as they finished the six month study period of the clinical trial (see Figure 2.1). In Study 1 (March to August 2022), we conducted focus groups and specifically exposed older adults in our target population to both an Echo Show device with Alexa and screenshots of EngAGE to help answer our research questions about perceptions of using digital voice assistants for preventative health. In Study 2 (May 2023 to September 2024), we conducted individual interviews with older adult participants immediately following their six months in the trial.

Through our focus group interviews for Study 1, we found that participants had a wide range of health conditions that influenced their attitude toward exercise. We report this here rather than in our findings since they are not research results but helpful context. Some explained that exercise could be difficult due to existing conditions (e.g., a prior stroke, fall, or car accident) and could potentially lead to injury. Other participants faced financial challenges (e.g., job loss) that often led to mental health challenges (e.g., depression). Many of our participants also struggled with getting out of the house, either to public spaces (e.g., a park) or exercise classes and these challenges were made even more difficult by the COVID-19 pandemic. Exercise was less challenging for some participants who noticed concrete benefits from weight loss while others said they had a lifelong motivation to be active. While participants in this study do share some characteristics (e.g., race, age, and neighborhoods), their relationships with exercise and aging are nuanced and varied.

9. The clinical trial is ongoing, so we are not able report medical outcomes in this chapter.

Table 2.1: Participant Demographics for Study 1

ID	Gender	Age	Race/Ethnicity	Education	Income[4]	Occupation	Session[5]
1	F	75-79	Black or African American	Associate's/Some College	N/A	N/A	1
2[1]	F	<65	Black or African American	Associate's/Some College	N/A	N/A	1
3	F	85-89	Black or African American	Bachelor's degree	$60,000/year	Retired	2
4	F	75-79	Black or African American	Associate's/Some College	N/A	Retired	2
5	F	70-74	Black or African American	Bachelor's degree	N/A	Retired	3
6	F	N/A[2]	Black or African American	Master's degree	N/A	Retired	4
7	F	90-94	Black or African American	High school degree	N/A	Retired	5
8[1]	F	65-69	Mixed race[3]	Associate's/Some College	N/A	Retired sales consultant	5
9	F	70-74	Black or African American	Associate's/Some College	$1,200/month	Retired	5
10	F	75-79	Black or African American	Doctoral degree	N/A	Retired pastor	5
11[1]	F	75-79	N/A	Associate's/Some College	N/A	Retired	6
12[1]	M	75-79	Prefer to self-describe: Mixed race	N/A	N/A	Retired IT specialist	6
13	F	75-79	Black or African American	Bachelor's degree	$70,000/year	Retired teacher	7
14	F	75-79	Black or African American	Associate's/Some College	N/A	Retired sales, insurance	8
15	M	70-74	Black or African American	Associate's/Some College	N/A	Retired	9

1: Denotes participant is an older adult caregiver. Participants $\geq$ 65 years old who identified as caregivers are also considered older adults.
2: N/A indicates participant selected: Prefer not to answer
3: Participant selected the following: American Indian or Alaska Native, Black or African American, White
4: Household income
5: Participants with the same Session number were in the same focus group.

2.3.2 Study 1

Protocol and Recruitment

Our focus group protocol was designed to answer our research questions regarding older adults' attitudes toward using a voice assistant for preventative health (RQ1) as viewed through a sociotechnical lens. To this end, the interview protocol included the following components: interview questions, an interaction with an Amazon Echo Show device, reviewing screenshots of EngAGE, and a sketching exercise. The interview questions focused on participants' social relations (e.g., *How do you usually communicate with your friends and family?*), previous technology use (e.g., *What type of technology do you use?*), and health-related activities (e.g., *How do you generally keep physically active?*) in order to understand the social context in which they learn about health and technology as well as their existing health and technology practices. This way of structuring questions has been shown to be effective in prior work on voice assistants for older adults [44]. We sought to structure interview questions to be asset-based and not presume any technology or health deficit in participants. The interaction exercise involved having participants read an overview of what Amazon's voice assistant does, set up the Amazon Echo Show device, and interact with Alexa by asking sample questions from a provided "tip sheet" (e.g., *Alexa, what's the weather today?*). The last part of the interaction exercise was having participants review screenshots of EngAGE.[10] The focus group ended with a sketching exercise that asked participants to draw where in their home they envisioned placing the Amazon Echo Show and how they pictured Alexa. Our sketching exercise builds on prior work that asked participants to draw their home networks, e.g., [145], and other work that asked participants to draw their relationship to technology, e.g., [176, 61, 63], in particular Melanie Hoff's project to have children draw Alexa [93]. In the single remote interview, we adjusted our interaction and sketching

10. Note, EngAGE was in the process of being updated at the time of Study 1 and thus unavailable for live testing.

exercises to send videos of someone interacting with Alexa and trying out EngAGE and had the participant create drawings at home that we captured via screenshot. Using an iterative approach, we updated the interview questions slightly over the course of the focus groups to better target our research questions.

We recruited both minoritized older adults (65+ years old) and their caregivers of any age who spent at least one hour a week with an older adult. In order to recruit participants, we placed flyers in a senior clinic[11] attended by people in our target demographic in a lower socioeconomic neighborhood and reached out to patients who participated in a previous geriatric study unrelated to voice assistants. Initial recruiting for the prior study was done by contacting eligible clinic patients directly; placing flyers in senior buildings, bingo events, grocery stores, and barber and beauty shops in the neighborhood surrounding the clinic; and snowball sampling. Between March and August 2022, we conducted 9 focus groups of 1-4 people that each lasted approximately 1 hour and were held in-person in a conference room at the clinic, with the exception of one participant who asked to do a video call over FaceTime. See Table 4.1 for a full list of participants as well as which focus group session they were in. All focus groups were audio recorded. Participants were each given a $25 Target gift card as compensation. We had interviews transcribed by Rev.com using a non-disclosure agreement.

11. The medical team has a longstanding relationship with a senior clinic that serves adults who are 65+ years old.

Table 2.2: Codebook for Study 1

Code	Code Description
Communication with friends & family	Current social relations and technologies used to keep in touch
Communication with doctor	Technologies used to talk to doctor
Physical preventative health	Physical activity, nutrition, source of preventative health info
Mental preventative health	Activities to stay mentally active
Medical concerns	Non-preventative health concerns and information sources
Current tech use & experience	Current devices used and frequency of use
Help with tech	Who helps set up and answer questions about technology
Previous voice assistant experiences	Prior interactions with voice assistants
Privacy concerns[1]	Any privacy concerns regarding voice assistants
Feelings	How did participants feel about interacting with Alexa
Feelings of friends and family	Friends and family feelings about voice assistants (perceived or actual)
Non-use	Specific tasks participants would not use a voice assistant for
To what extent use in future	In general, perception of future use
Relationship to Alexa	Characterizations of Alexa and participant relationship to the personality/device
Feelings about exercise application	Specific thoughts about the EngAGE program
Interaction breakdowns with Alexa	Examples of breakdowns when participants try out Alexa during focus group
Location of voice assistant at home[2]	Where place voice assistant in the home
Outlook on life/aging	General thoughts on life, aging, philosophy, etc. that impact views on technology
Mental models[3]	Mental models of voice assistants

1: This chapter focuses on analysis of the bolded codes.

2: Sketches from the prompt "Draw where you would put the Echo Show device in your home" analyzed under this code

3: Sketches from the prompt "Draw what you think Alexa looks like" analyzed under this code.

Participants

We recruited N=15 total participants who are listed in Table 4.1. Participants included 14 minoritized older adults, three of whom identified as caregivers as well, and one younger caregiver. Most of our participants were women older than 75, retired, self-identified as Black or African American, and had at least an Associate's degree or some college. Many participants chose not to report annual household income. The median household income (2016–2020) for the neighborhood in Chicago where we recruited was $35,887/year. Approximately 82% of residents in this neighborhood have some form of Internet access (2016–2020).

Analysis

Data analysis was completed in two rounds: an initial coding phase and a second round of coding for a thematic analysis [162]. In the first round, we collaboratively developed the codebook in Table 2.2 based on our focus group protocol and research questions. Two members of the research team then coded all transcripts using the codebook in the qualitative software program MAXQDA. The two versions of the codes were merged in MAXQDA (since we merged the codes, we did not calculate inter-rater reliability [121]). In this phase, we also analyzed the sketches from participants (i.e., the graphic representations) using two codes: 1) sketches from the prompt *"Draw where you would put the Echo Show device in your home"* were included in code *"Location of voice assistant at home"*, and 2) sketches from the prompt *"Draw what you think Alexa looks like"* were included in code *"Mental models."* In the second round of coding, one member of the research team coded for subcodes in each top-level code. After both rounds of coding, each team member produced a thematic summary of the top-level code and its subcodes. Several members of the research team reviewed the thematic summaries and discussed the final emergent themes to include in this chapter. In this chapter, we focus only on a subset of codes bolded in Table 2.2 because they most closely informed our research questions.

Table 2.3: Participant Demographics for Study 2

ID	Clinical trial group	Gender	Age	Race/Ethnicity	Education	Indiv. Income (year)	Occupation
16	EngAGE	F	70-74	Black or African American	Master's degree	$0	Retired
17	EngAGE	F	65-69	Black or African American	Associate degree	$40,000	Retired
18	EngAGE	F	70-74	Black or African American	Bachelor's degree	$60,000	Retired
3[1]	EngAGE	F	85-89	Black or African American	Bachelor's degree	$50,000	Retired
20	EngAGE	F	70-74	Black or African American	Bachelor's degree	$24,000	Retired
21	Paper	M	65-69	Black or African American	Some high school	$26,400	N/A
22	EngAGE	F	75-79	Black or African American	Some middle school	$10,968	Retired
23	Paper	F	80-84	Black or African American	Bachelor's degree	$75,000	Retired
24	Paper	F	65-69	Black or African American	Some college	$13,000	N/A
25	EngAGE	F	65-69	Black or African American	Some college	$0	N/A
26	Paper	F	65-69	Black or African American	Some college	$30,000	Retired
27	EngAGE	F	70-74	Black or African American	Some high school	$16,800	Retired
28	Paper	M	65-69	Black or African American	Some college	$84,000	Retired
29	Paper	F	80-84	Black or African American	Master's degree	$0	Retired
30	EngAGE	F	70-74	Black or African American	High school	$16,668	Retired
31	Paper	F	65-69	Black or African American	Some college	$10,020	Retired
32	Paper	M	70-74	Black or African American	Associate degree	$14,400	Retired
33	Paper	F	65-69	Black or African American	Some high school	$10,968	Retired
34	Paper	F	65-69	Black or African American	Some high school	$10,092	N/A
35	EngAGE	F	65-69	Black or African American	Some college	$20,000	Retired

1: P3 was in both Study 1 and Study 2.

2.3.3 Study 2

Protocol and Recruitment

We started conducting interviews for Study 2 in June of 2023, as participants began finishing the six month clinical trial (as described in Section 2.3.1 and shown in Figure 2.1). Given that Study 2 involved two separate clinical trial groups (paper exercises and VA-based EngAGE), we developed two parallel interview protocols to identify the primary motivating factors influencing Black older adults' decision to exercise, and what role a voice assistant played in this process (RQ2). Both protocols had questions about a participant's previous experience with exercise, their routine with EngAGE or the paper packet, their social support mechanisms (e.g., *Who in your life helps you to stay healthy?*), their future exercise habits (e.g., *Do you think the paper/EngAGE exercises have changed how healthy you are?*), and their interactions with the Alexa device, which participants in both arms received (e.g., *Did you ever use [Amazon Alexa]?*). In the interview guide for the EngAGE group participants, we asked additional questions about features that only existed on the EngAGE (e.g., *Did you use either the reminder or tracking system?*) and about interactions that are only possible with a VA device (e.g., *Is there anything you wouldn't say with Alexa in the room?*).

As shown in Figure 2.1, we did not conduct the original recruitment of participants into the clinical trial, which was done by the clinical trial team. The clinical trial screened for older adults over the age of 65 who were community-dwelling (i.e., not living in hospitals or nursing homes, but could be in seniors-only housing), Black or African-American, living with two or more long-term illnesses, and who had difficulty leaving home. We conducted our interviews for Study 2 immediately following the end of a participant's six month period in the clinical trial (see Figure 2.1). These interviews were conducted over the phone, lasted no longer than 30 minutes, were audio recorded, and occurred from May 2023 to September

2024.[12] As in Study 1, we updated the interview questions slightly over the course of the interviews to better target our research questions and had the interviews transcribed by Rev.com using a non-disclosure agreement.

Participants

We recruited N=20 participants for Study 2 from the pool of participants in the clinical trial (see Table 2.3). At the time of our interviews, Study 2 participants had been using the exercise materials via EngAGE (N=10) or the paper packet (N=10) for six months (see Section 2.3.1). All but six of the Study 2 participants had some higher education; they were predominantly retired women in their 60's and 70's whose individual income ranged from $0 to $84,000 per year, though we note that 15/20 participants made under $35,000 per year. Participants had to identify as African American or Black to be eligible to participate in the clinical trial. The demographics in Table 2.3 were collected by the clinical trial team, which is why some demographics are reported differently. Note there is one participant, P3, who participated in Study 1 and Study 2.

Analysis

We analyzed the interview transcripts using two rounds of coding for a thematic analysis with a team of three researchers similar to the process used in Study 1 [162]. First, two members of the research team developed and refined our codebook, based on the interview protocol, by each coding three of the same transcripts and ensuring agreement in codes (see the codebook in Table 2.4). Then, three members of the research team used MAXQDA to each code one third of the transcripts using this codebook. Each coded transcript was then coded in a second round by a second coder from the team (Since we used the coded

12. The long time period is a result of the significant challenges of recruiting participants in this population. Every participant was in the trial for six months total, but start times varied.

Table 2.4: Codebook for Study 2

Code	Code Description
Prior exercise habits	Types of exercise the participant engaged in prior to the study
Motivation to exercise	Reasons for why the participant felt motivated to exercise and stay healthy
Challenges to exercise	Reasons why it was challenging for the participant to exercise
Routine	What kind of routine the participant had for using the EngAGE or paper exercises
Feedback on study exercises	Comments related specifically to the exercises/content provided in the study
Feedback on study	Miscellaneous study feedback not related to the exercises/content
Care partner relationship	What was the relationship of the study-designated care partner to the participant
Care partner support	The kinds of support (or lack of support) provided by the study-designated care partner
Non-care partner support	The kinds of support provided by someone other than the study-designated care partner
Future exercise plans	How does the participant envision their exercise routine going forward
Voice assistant uses	Uses for voice assistants (including Alexa) outside of the exercise program
Voice assistant opinions	Opinions about voice assistants (including Alexa) outside of the exercise program
Other motivating factors	Motivating factors not captured in existing codes

data as input to thematic analysis, we did not calculate inter-rater reliability [121]). During the coding process, the entire research team met regularly to discuss the codes and ensure agreement. Using the coded segments, we collaboratively wrote thematic summaries and then based on these summaries and discussions, we identified a list of factors motivating participants to exercise and grouped these into interaction levels, which we summarize in Table 2.5.

2.4 Study 1 Findings: Perceptions of the Alexa Voice Assistant

In this section, we present our findings from the first study and address RQ1 by describing participants' perspectives on voice assistants (VAs) based on their interaction with the Amazon Alexa VA during our focus groups. We found that our participants felt a wide range of different relationships with the VA, that for some shifted throughout the focus group as they grappled with how to make sense of the device. The way that participants framed this relationship impacted how they interacted with the device and how they envisioned using it for preventative health. The context in which participants were using, or imagined using, a VA also informed the type of relationship that was foregrounded between a participant and the VA (e.g., P11 characterized the VA as a *"peeping tom"* when discussing its voice detection and recording and a *"beautiful woman"* when describing the interface and advanced technological capabilities). In what follows, we sort our findings into distinct relations participants' described with the Alexa personality and describe how these impacted participants envisioned use of the device for preventative health applications.

2.4.1 Alexa as Tool

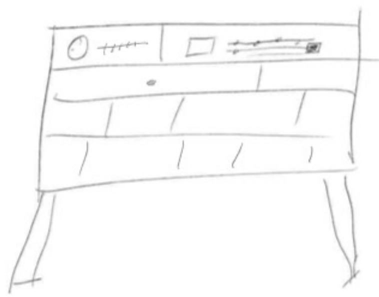
Some participants saw the VA as a means to an end, or a **tool**, for accomplishing preset goals or tasks. This conceptualization aligned with participants who envisioned the VA as a machine, as seen in the first drawing in Figure 2.2. In this section, we describe the uses related to promoting convenience and utility in participants' day-to-day, such as using the VA as a timer, alarm clock, video-calling device, voice-to-text/text-to-voice, and search engine. Participants explicitly related several of these uses to exercise, diet, and other health applications.

Participants who wanted to use the VA for exercise were interested in the VAs ability to help them **form and keep up exercise habits**. P4, for example, often found herself letting herself *"go by the wayside,"* by *"playing Wordscapes"* first thing after waking up, rather than doing her morning stretches. P3 shared a similar struggle, but imagined the VA could catch her *"between the time [she] open [her] eyes and before [she] jump[s] out of bed"* by reminding her what exercises she should *"do before leaving the bed."* P5 thought it would be *"encouraging for Alexa to remind"* her to go on her daily morning walk and P15 hoped he could use it as a *"home assistant"* that could work as a *"timer for high intensity exercise."* Maintaining an exercise schedule structure was an important unmet need for many participants that they imagined could be filled by the VA.

The ability to **search for information** on the VA attracted participants who were interested in using the VA for their health. Exercise instructions were one of the types of information participants searched for. It could be an easy way to learn exercises for those unable to easily leave their homes, even if just temporarily, such as P5 who reflected that, if she owned a VA during bad weather, could *"tell [the VA] what [exercises she is] looking for"* and *"could have been inside doing"* them. Others thought to search for recipes to help *"track and maintain"* food intake under *"diet care constraints"* (P15). Some participants envisioned looking up non-urgent health and medical questions. For example, P10 wanted to use use the VA to do *"research on [her] own"* for checking information about her medication. Other health applications included emergency and non-emergency hands-free communication with medical professionals (P2, P14) and using brain teasers to keep the mind active (P11). In the specific context of retrieving and learning information (see Section 2.4.5 for lack of trust contexts), participants found the VA reliable and trustworthy, making it a good tool for health and exercise education.

2.4.2 Alexa as Amusement

Participants were used to using technology as a way to access amusement, be that via their cable televisions, smartphones, music players, or DVD players, and saw a VA fulfilling a similar role by **integrating it with the technologies they already use**. P5, for example, thought of leveraging vocal commands on the VA to use it as a remote control, or to *"turn the TV off"* (P4). P7, who watched church services on her tablet, thought to do the same on the VA device, as suggested by P8 (her caregiver). P3, while not sure if the VA had this capability, said she would like to use the device to *"play [music] that's on [her] phone."* These examples, while not directly related to health, show that participants were willing to



(a) P11's sketch of Alexa as a machine.

(b) P15's sketch of Alexa as a woman waving.

Figure 2.2: Participant sketches in response to the prompt "Draw what you think Alexa looks like."

incorporate the VA into existing daily interests and habits, which could include exercise.

For several participants, a gadget like the the VA device was akin to a **toy or game**. P6, who was especially excited about the device's screen, said she could *"get lost at this thing playing with it."* Playing meant different things for different participants. P4, for example, would test the VA's limits by asking *"stupid questions,"* like *"Alexa, where my car at?"* For P11, the entertainment came from the games you could access through the VA (e.g., *"crossword puzzles"*). She stressed the need to *"exercise the brain"* by using its *"muscles."* Participants' interest in using a VA for improving their memory and challenging their brain shows that they can envision using the entertainment-focused aspects of the technology for the benefit of their health.

2.4.3 Alexa as Friend, Lover, and Family Member

While some participants' real and imagined uses of the VA were primarily task-oriented (e.g., a time keeper in Section 2.4.1 or toy in Section 2.4.2), in other cases the participant-VA relationship emerged as the most influential aspect to the interactions and imagined uses. Figure 2.2 exemplifies this contrast in views of Alexa as a static object (a machine in the shape of a box, in P11's sketch) and Alexa as a human or human-like being (a woman waving, in P15's sketch). Participants would sometimes refer to the VA or interact with it as one would in a **friendship, family, or romantic relationship**. P3, for example, called Alexa her *"good friend"* and P6 said the VA she owned before the study was *"kind of cool to hang around."* For P14, the *"awesome"* relationship was even more significant since she is *"a person living alone, having no contacts, doesn't do much and doesn't see a lot of people."* Some of these relationships even crossed into romantic territory, though mostly in a joking way, such as for P15, who was considering whether he would adopt the study's VA in addition to his existing VA, Siri, and remarked he was *"trying to be monogamous as much as possible"* when it came to VA technology. The companionship evident in these comments and interactions reveal that participants can envision integrating a VA into their lives in an important role reserved for intimate relationships.

While participants who spoke of the VA as a friend or romantic companion often did so in a positive tone, participants' attitudes toward a VA taking on a **familial role** were more complex. In some cases, the VA was a parental figure: *"[Alexa is] your mommy getting after you as a little kid and she got a good mommy voice"* (P3). Other times, the participant was the parent. Many participants referred to Alexa as a *"girl"* and did not seem to think it was particularly smart. P1, for example, remarked that the VA was *"acting like a five year-old"* when it kept speaking even after being asked to stop. P5, similarly, was not successful in asking more complicated questions to the VA, concluding that, to successfully communicate with it, *"you should just be straight with her."* The infantilization of the VA meant some participants approached it with tenderness too. P5, after her *"straight"* talk comment, called the VA *"cute"* and P3 spoke to the special place her own VA had in her life as one of three members of her inner circle: *"that's my husband and this girl."* The familial role of the VA is more nuanced than the friend or romantic role since it introduces power dynamics that often call into question the VAs ability and competence.

2.4.4 Alexa as Stranger or "Other"

Some participants found Alexa **unrelatable**; this was most evident during the live interactions between participants and the device during the focus groups. P3 and P4, for example, asked Alexa for *jerk chicken* recipes and while it produced recipes, none were specific enough since *"she doesn't know jerk"* (in this context, "jerk" is a Jamaican style of cooking). P15 found it *"very frustrating"* trying to communicate with Alexa not because of what he was saying, but because of how he was saying it. *"They not really set up for people of my ethnicity,"* he remarked, *"it doesn't recognize the tonality or whatever it is."* All three participants, therefore, observed that Alexa did not understand their requests, implying that perhaps the device was not indented for people with their cultural background.

Other participants encountered this cultural barrier and attempted to troubleshoot the issue, or at least diagnose it, but were unsuccessful. For example, when P5 asked Alexa to tell her a joke, the VA responded with: *"Why did the golfers get kicked out of the library? Their pants were way too loud."* This joke, which would only be accessible for an audience familiar with golfing culture to some degree, left the participant puzzled and wanting to know more about the background of the designers of the VA to see if that might explain the joke. When P5 asked Alexa this exact question, it responded with an unsatisfying joke answer that did not reveal any new information to her. These barriers were noticeable to participants and eroded trust in the relationship. In fact, it made some participants, such as P15, disinterested in using the VA in any context, including for exercise or health promotion.

2.4.5 Alexa as Intruder

Some participants described fundamental issues they had with the VA device that stopped them from forming any positive relationship with it; these participants saw Alexa as ***"an intruder"*** (P1). P14, for example, had moral objections: *"The AI thing, I hate it. Just something about it. It's so unnatural that we're doing too many things that God did not*

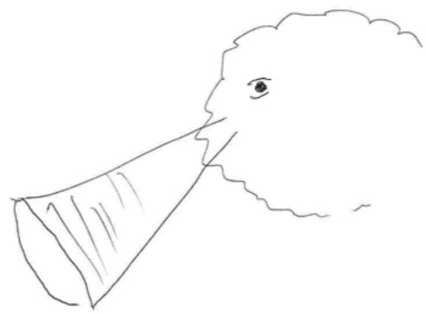


Figure 2.3: P12's sketch of a cloud looking in a telescope for the prompt "Draw what you think Alexa looks like" that he labeled *Spy in the Sky* to show that he felt cloud-based Amazon Alexa devices were spying on users

intend for us to do." Another key issue was privacy. P8 said that while her own activities were *"not that important"* to warrant the VA or its designers tracking her, she had no doubts that the manufacturer *"could see what everybody's doing."* P11 and P12, a married couple, feared being hacked and surveilled by the VA device and other technology. P12, for example, noted that he chose to not purchase a smartphone *"because it can be hacked"* and, as a result, was uncomfortable *"putting [his] whole life"* on it. They shared similar reservations about owning a VA device, calling it a *"spy in the sky,"* illustrated in P12's sketch in Figure 2.3, due to its ability to monitor its users. P1, similar to other participants, was not concerned about spying from the VA manufacturer, but from his own family, saying his data *"could be unloaded by any grandchild, which makes me feel very icky."*

The degree to which participants believed they were in control had some impact on their willingness to use the device. P1, for example, presented the following strategy for limiting the device's access: *"If I was doing something, I'd have to take a brown paper bag, open it up, turn it upside down, and put it over Alexa for my own privacy."* P14 similarly understood physical space to play a part in how much privacy mattered, noting, *"If I had it in that one room upstairs, I'm only in there when I read, meditate, pray, take my... And then I'm out of there."* Ultimately, though, the threat of surveillance was unacceptable for some participants, even those who presented strategies for mitigating privacy loss. P1 *"would not allow this in* *her house"* and P15 was *"not going to go out and buy [it], for [it] to come in and spy on [him]."* In the most extreme case, P12 chose not interact at all with the Echo Show, stating *"I'm not recording my voice."* When participants spoke of Alexa as an intruder they were the clearest, compared to any other relationship, to point out that the role the VA would occupy in their lives was unacceptable and, therefore, they were not willing to use it for exercise, health, or any other use.

To summarize Study 1, participants identified moral and cultural barriers and expressed complex attitudes toward privacy loss from using voice assistants. While participants were frustrated when facing the former, they never gave up using the device because of cultural barriers. Some participants were similarly able to overlook or address their privacy concerns with the device, but for others the potential privacy loss outweighed any possible benefit gained by using a voice assistant to promote preventative health.

2.5 Study 2 Findings: Alexa Voice Assistant vs. Paper Packet Exercise Routines

In this section we address RQ2 to understand how older adults perceived a voice assistant exercise program versus the same exercises provided in a paper packet. To do this, we compare factors influencing participants to exercise in each group. We organize these motivating factors into four interaction levels: 1) individual, 2) human-artifact interaction, 3) social support interaction, and 4) research study interaction (see Figure 2.4 and Table 2.5). At the *individual* level are motivating factors that are unique to the person and their circumstances, such as their current health status. At the *human-artifact interaction* level are motivating factors that are derived from the exercise artifact (VA or paper packet), such as Alexa's personality. At the *social support interaction* level are factors related to ways friends and family motivated participants. At the *research study interaction* level are factors related to the motivation found from the clinical trial structure and staff. We sort the motivating

Table 2.5: Study 2: Motivating factors that influence a participant's exercise behavior

Interaction level	Factor	Description
1. Individual	- Intrinsic motivation to exercise - Lack of health challenges	- Personal belief that exercise is important - Participant in good physical state
2. Human-Artifact	- Exercises feel achievable - Noticeable health-related improvements - Voice assistant device interactivity - Alexa's personality - Voice assistant privacy	- Exercises are not too challenging for participant - Participant felt better after doing exercises - Interactivity affordances such as voice commands - How participants felt about interacting with Alexa - Concerns about the privacy and security of VA data
3. Social Support	Care partner support - Broader social support	- Support provided by designated care partner in study - Support provided outside of designated care partner
4. Research Study	- Habit formation due to clinical trial structure - Technical support and comfort with technology	- Exercise became a habit during clinical trial -Technical obstacles overcome due to tech support staff

factors into these levels (see Table 2.5) as a way to organize our findings and underscore how the entire sociotechnical system was important for motivating study participants to exercise, not just the exercise artifact in isolation.

Overall, we found that most influential motivating factors—such as social support and lack of health challenges—were present for both the voice assistant and paper study groups. The VA itself was a positive motivating factor for some, but a negative motivating factor for others. This finding highlights again what we found in Study 1: that participants' perceptions about the device and Alexa's personality influence their willingness to use the VA for health purposes. For those that had positive experiences with the VA, it was a tool to *amplify* motivation to exercise but could not replace the other factors in the sociotechnical system required to support older adults in adopting an exercise routing.

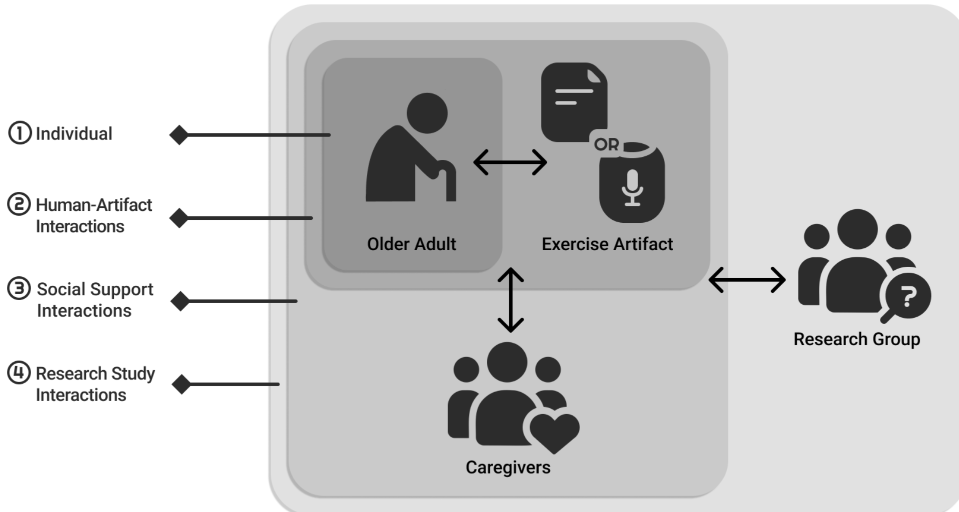


Figure 2.4: Interaction levels experienced by clinical trial participants

2.5.1 Individual Factors Influencing Exercise

Individual motivating factors were characteristics of the participant that made them more likely to exercise (1 in Figure 2.4). We found both factors—intrinsic motivation and lack of health challenges—were important in both the VA and paper packet groups. Access to the VA device did not appear to influence either individual motivating factor.

Intrinsic Motivation to Exercise

Multiple participants in both the VA and paper exercise groups described that they had always been self-motivated to exercise, which we call intrinsic motivation. We found this was an important factor in whether people exercised. P3, in the VA group, said *"Well, see, I've always told myself from a young woman that I'm total, whole, complete, unlimited, and highly adaptable. So, your body will give you back that if that's what you tell it."* A representative comment from the paper exercise group was P28 who said, *"I tell somebody if they want to get into the program to really get themselves together, self-motivation is the key to anything."* Both groups felt self-motivation was key to sticking with an exercise routine.

Lack of Health Challenges

Another important factor influencing whether participants in both groups exercised was a lack of health challenges. Even those with intrinsic motivation could be stymied by setbacks such as surgery, arthritis, bad backs or knees, or chronic pain due to prior injury. In the VA group, P16 said *"Since last September and all through [the clinical trial], I've had some health issues which got in the way of my doing the EngAGE and other exercises."* In the paper group, P21 said *"Because I was in so much pain. And I mean, my knees, my back, you name it. I was in so much pain and it was hard to do it."* Participants in both groups might not exercise due to health challenges.

2.5.2 Human-Artifact Factors Influencing Exercise

In this section, we explain motivating factors derived from interaction with the artifact (either the Amazon Echo Show with the EngAGE or a paper packet with the same exercises; 2 in Figure 2.4). Some motivating factors related to the exercises themselves were found in both groups. Specific to VA interaction, interactivity and Alexa's personality could be positive motivating factors while Alexa's personality and privacy concerns could also be negative factors. Overall, we note that our sample of participants was heterogeneous on whether or not they enjoyed interacting with a VA, building on our findings from Study 1.

Exercises Feel Achievable

Participants in both groups described that part of why they continued with the exercise program was because the exercises felt achievable and within their physical capacity to complete. P27, in the VA group, said *"It wasn't strenuous or none of that. It was something that I could do. So I actually liked it."* Similarly, P21 in the paper group said he was initially were concerned it would be too difficult *"at this age and out of shape, I'm already in pain because the arthritis and all that"* but *"when I did (the exercises), I was amazed. I was like,* *'Wow, that is real smooth."'* Thus participants in both groups were motivated to exercise if they felt the exercises themselves were not too difficult or strenuous. One affordance offered by the EngAGE was the ability to verbally respond if an exercise was too easy or too hard. The paper packet had notes on how to vary the difficulty, but was not interactive like the VA app. P3, in the VA group, noted that *"When I first started out... I answered [the exercises] were too difficult. And as I progressed... it got easier, so I was always just right."* None of the paper group participants spoke to the ability to vary the difficulty of the exercises, suggesting that the VA can enhance this motivating factor.

Noticeable Health-Related Improvements

Participants in both groups said they were motivated by seeing or feeling health-related improvements over time that they attributed to the exercises. The most common improvements we heard about were weight loss and better mobility. P30, in the VA group, was motivated by weight loss saying, *"I'm glad I started, because I did lose some weight, and I was very proud of myself. I was like, 'I actually want to get this exercise done."'* P23 in the paper group was motivated by mobility saying, *"Because when I don't exercise, I get really, really stiff, and my arthritis acts up more."* P21 in the paper group simply said, *"I felt a lot better afterwards. You know? It made me be able to get up and move around a lot easier than I would if I hadn't have done it."* Interestingly, the exercises were not designed for weight loss. Several participants reported that they also paid more attention to nutrition during the study period (again this was not prompted by the study design). It may be that in some cases a regular exercise routine leads to positive outcomes such as a healthier diet. Overall, participants were more motivated to keep exercising if they attributed noticeable, concrete health benefits to it.

Voice Assistant Device Interactivity

Within the VA group, some participants found the interactive affordances of the VA device—such as the voice prompts and screen images—helpful. P18 provided a strong endorsement saying, *"I liked it, the screen showing you what you're doing at a particular time, and then Alexa reminding you how to breathe because I forget that, when you're supposed to breathe in and when you're supposed to breathe out. I get confused and then her talking to you brings it back to me."* We also found that most in the VA group spent 30 minutes exercising, which was the default for the program. For example, P30 said, *"I just follow along with the instructions that [Alexa] tell me what I have to do and I just do it."* Interactivity and pre-set default exercises can help motivate participants to achieve a particular level of exercise desired by the designer (as long as they feel achievable, as we saw in a prior section).

Alexa's Personality as Positive Factor for Some, Negative for Others

Like we found in Study 1, participants had mixed opinions on Alexa's personality, which was important because most considered Alexa and EngAGE as one experience. Some participants found Alexa amusing, for example, P16 said, *"And it is funny because if I whisper, she whispers right back at me."* P19 was lukewarm on Alexa's personality, saying their interaction was "fine" but said that she looked forward to learning how to use a new technology: *"I know how to use the television, I know how to use Zoom. My brain is still wanting to learn. So, if there was something else given to me, I'm very anxious to learn something."* In this case, the participant's excitement about learning compensated for her neutral opinion of Alexa. Alexa's personality was a negative motivating factor for others. For example, P25 said, *"I wasn't crazy about EngAGE... I rather somebody just give me the exercises to do and then just let me do them rather than her talking all the time because that's a distraction and an irritant for me."* This was the strongest negative comment we received, however, others were neutral or lukewarm on their interaction with Alexa. Perceptions of Alexa's personality impacted the extent to which participants were motivated to use the VA device for exercise.

VA Privacy Concerns as Negative Factor for Some, Neutral for Others

In addition, numerous participants were concerned about the privacy implications of having a VA device in their home. P27 made a representative comment saying, *"[Alexa] hear things that she ain't got no business hearing. Then, she might repeat something, because somebody said something, but not necessarily talking to her... If there's something going on, and the police is called, she going to tell everything she knows."* In contrast, other participants were not concerned about privacy risks from using the VA device or felt they had nothing to hide. The level of privacy concern about the VA device impacted if participants wanted to use it for health applications including exercise. As we saw in Study 1, some were totally unwilling to have a VA device in their home, and these participants would not have signed up for the clinical trial at all.

2.5.3 Social Support Factors Influencing Exercise

Social support includes the encouragement and and assistance friends and family provided to help the participants stick to their exercise routine (3 in Figure 2.4). As part of the clinical trial, every older adult participant chose a "care partner" to support them through the study period. We found chosen care partners were almost entirely family members, consisting mostly of children and grandchildren of the participant. A couple of care partners were neighbors. We found that care partners were critical exercise motivators for participants in both groups, and many also received additional support from their community.

Study Care Partner Support

In both the paper and VA groups, participants reported that care partner support was a major factor in keeping them on track and motivated to exercise. For example, P26 (paper group) said, *"(My care partner) kept me dedicated to (the exercise program) because I knew he would ask me about it. So I didn't want to lie, so I would just go ahead and do it."* In another representative quote, P29 (paper group) said, *"(My care partner) calls me, she reminds me, and at least three to four times a week she'll come over and partner up with me and do the exercises."* In one case a care partner for P33 (paper group) moved away mid-study and the participant said that after he left *"The motivation wasn't there anymore."* Having a trusted partner was important: P25 (VA group) said, *"Because (my care partner) knows me and it wasn't like it was a stranger or someone distant. When she says things to me, she wants me to do better."* P30 (VA group) had her son as her care partner and explained, *"He was just always asking me, 'Mama, how did the exercise go or did you enjoy doing it?' I just say, 'Yeah, it was fun."'* Our findings show that while the VA group was able to interact with a conversational human-like agent (Alexa), this did not replace the need for a human friend or family member to offer support and motivation.

Broader Social Support

Participants also described the importance of social support for health outside of the designated care partners, with close family and friends again being the most critical. Several participants mentioned groups of friends who would walk together, or seeing friends at the gym, as reasons for staying motivated to exercise. Participants also explained how family would help them with nutrition and a healthy diet. In Study 1, some participants also mentioned that before the COVID-19 pandemic, they liked going to exercise classes at a community center but had stopped exercising entirely once they had to stay at home. This demonstrates how the VA can augment an exercise routine, for example if a participant cannot make it to the gym, but that community-based activity is still a strong motivating factor that the VA cannot replace.

2.5.4 Research Study Factors Influencing Exercise

In this section, we examine ways in which the clinical trial itself provided motivating factors for participants to exercise, including the structure of the clinical trial and the staff tech support (4 in Figure 2.4).

Habit Formation due to Clinical Trial Structure

The clinical trial provided VA and paper group participants with a structured exercise routine—including the exercise artifact, the six month timeline with regular check-ins from study staff, and instructional materials—that led them to develop new exercise habits. P31 (paper group) had a representative comment to this end saying, *"I hadn't done any surveys or studies or anything like that in a while, and I was pretty much just going along, not doing whatever I wanted to do. But then this gave me some structured time to do things."* P31 continued saying that she would keep up the exercises because *"I don't think that you can do something like this for such a length of time and then all of a sudden just stop it."* In the VA group, P20 also mentioned the importance of a structured routine saying, *"It was good to have that routine... I know I got to do Alexas and then if I miss, I said, 'Oh, I didn't do my Alexas yet. I got to do them by a certain time."'* While the clinical trial did not provide any hard requirements around doing the exercises, the structure of the trial was a motivating factor in participants developing an exercise habit for both groups.

Technical Support and Comfort with Technology

Participants' comfort level using the VA device was significantly bolstered by help from a full-time clinical trial staff member who acted as on-call technical support for participants. At the end of the trial period, participants lost this constant technical support and in addition needed to transfer the VA to their personal Amazon account (not the clinical trial account). This created a barrier for continued use for some but not all participants. P16 was unable to keep using EngAGE because she could not figure out how to transfer the device to her account saying, *"Because even after [the clinical trial account] was gone... I tried to, dang, I can't do this anymore. So I went to do [the exercise program], but I couldn't."* Similarly, P18 said *"I'm not knowledgeable of technology. So [technical support staff] was explaining some things to me, and I'll have to get my nephew to come over and set it up and explain it to me too... Because all I was doing was telling Alexa, 'Let's start the new EngAGE program."'* These participants wanted to continue using the exercise program, but the technical barriers were too great for them to overcome without help. In this case, paper exercises provide an advantage since motivated people can use them without aid.

2.6 Discussion

In Study 1, we organized participant perceptions of Alexa into relational types that ranged from positive (e.g., friend) to neutral (e.g., tool) to negative (e.g., intruder) and impacted participants' willingness to use the voice assistant (VA) for health. In Study 2, we compared motivating factors to exercise in participants who had spent six months using a VA exercise program versus those who had a paper packet with the same exercises. We found many motivating factors were shared between the groups; the VA could be an additional motivating factor for some but not all participants.

In this section, we situate these findings in work on "contextual factors" in health technology solutions for older adults. Contextual factors can include cultural aspects, social context, race/ethnicity, and access/affordability and these factors influence if and how a technology is adopted [154]. The way Black older adults describe how they relate to Alexa is important context for designing health VA applications because it is grounded in the history and culture of Black communities in the US (Section 2.6.1). The importance of integrating social support into VA preventative health applications also provides the social context needed for a successful intervention particularly in lower-income communities (Section 2.6.2). This work builds on the tradition of CSCW research on understanding how technology is embedded in social systems, and in particular on literature on interpersonal relations with VAs (e.g. [147, 163, 216, 80]) and social support in health technology interventions (e.g. [89]).

2.6.1 Designing participant–voice assistant relationships with Black older adults

How Black older adults relate to Alexa has implications for designing health applications that will be used and trusted by this group. Literature on interpersonal relations with VAs has identified a dichotomy between treating the VA as either a person or an object [147] and found that when VAs are anthropomorphized they are typically identified as a friend [147, 163, 216, 80]. Studies on interpersonal relations have not focused on racial/ethnic identities of older adults, however, prior work has investigated how voice assistants like Alexa have tended to be designed to sound white (not using or understanding African-American Vernacular English) and female [144, 166]. This has led Black users to explain how they "code switch" to change their style of speaking when talking with voice assistants [88, 31]. We find that while some Black older adults in our study thought of Alexa as a friend or an object, other categories emerged and not all were positive.

Some participants saw the VA as a *tool*, *friend*, and *amusement*: neutral to positive relationships. While prior work has focused on the friend category, we want to build on this to highlight amusement as an additional positive design lens for creating VA technology for older adults (of any race), even when the outcome (health) is not entertainment. Most of our participants were retired, and many were confined to their homes for extended periods of time due to physical limitations. Alexa provided an amusing distraction for many, as well as a way to learn a new skill. Prior work has shown that for some older adults, the cognitive burden of new technology can be a barrier to use [148, 217, 107]. We find this is true for some, but participants also reported enjoying the experience of exploring and learning a new technology to keep their brain sharp and often used it as a talking point with younger family members [193]. We suggest that designers of VA health applications leverage entertainment value that participants experienced when interacting with the VA or other technologies.

Another important relationship that influenced participants' willingness to use the VA in Study 1 was the *stranger*, highlighting cultural differences between the VA and the participants. This manifested, for example, when Alexa was unable to provide a recipe for jerk chicken, a traditionally Jamaican dish, or when the VA made a joke about golfing (a sport predominatley practiced by white, high-income people) that participants did not understand. Participants felt alienated and frustrated by these experiences; one even concluded he must not be the target audience for the VA given it struggled to understand him. Our findings build on prior work that suggests that some Black older adults do not have a positive view of VA devices, in part due to cultural and language differences, and suggest that making more VA personas might help [87]. As new AI methods improve the ability of conversational VAs to mimic a range of languages and cultural content, future work should investigate how the perceptions of Black older adults change when VAs can engage in more relevant dialog and content.

A final relationship is the *intruder*, based on some participants' concerns about privacy that have roots in the history of surveillance of Black people in the US. While previous work has examined older adults' trust and privacy attitudes toward technology [22, 77, 76, 82, 111, 60, 106, 94, 173] and how these may or may not impact voice assistant use [109], these studies were either completed with majority white participants or did not discuss their race/ethnicity. The conscious consideration of race as a factor informing our participants' trust and privacy attitudes toward voice assistants is particularly salient given prior work on surveillance in Black communities [32] and the ongoing incorporation of racial discrimination into new technologies [18]. The distrust of the older adults in our study was culturally and historically located, especially given that many participants have lived through the civil rights movement and the introduction of the Internet. These participants who were privacy-oriented, but still had an interest in using a VA or owned one, described using paper bags and physical separation to address their surveillance concerns. Designers of VAs could leverage physical or hardware based controls (e.g., camera switches or removable microphones) to provide strong privacy guarantees to participants who may not trust software-based privacy solutions. The CSCW and related communities have already begun to investigate more tangible privacy solutions for smart devices e.g., [7]. Further research should be done to understand how older adults from varying demographics (e.g., race/ethnicity, income, gender) conceptualize privacy and surveillance harms when using voice assistants to both address and mitigate these concerns.

In these suggestions we argue that how users relate to a VA device is culturally situated and taking these interpersonal relations into account in the design phase is critical. In our study, we saw that cultural and historical understandings of recipes, jokes, and privacy contributed to how comfortable Black older adults in the US felt about interacting with a VA. Considering these relationships is especially important in the stranger and intruder contexts since these were significant obstacles to our participants for using the VA in any context, including health. Rather than reconsidering the design of the health intervention at the health application level, it is necessary to perform a deeper analysis into the relationship between Black older adults for whom privacy and contextual understanding are important. Just as the VA made assumptions about the types of jokes that will land for participants and failed to consider non-hegemonic uses of language, it may, by default, make health and exercise suggestions that alienate some of its users, such as those with specific disabilities or those without access to maintained public parks. As reinforced by prior work [203], considering cultural and historical factors when analyzing privacy attitudes and behaviors is crucial to the development of technologies like voice assistants.

2.6.2 Human social support is a critical component of voice assistant preventative health application design for lower-income older adults

In Study 2, we teased out 11 factors that motivated participants to exercise and compared these factors between the VA group and the group with a paper packet of exercises. Prior work on technological physical health interventions demonstrated that older adults seek out and find signficant gratification in social connections [208, 199]. We found that in addition to being a key value, social support and connectedness can also bey a key motivator for exercise. For example, many participants said that their designated care partner checking in on them, and in some cases doing the exercises with them, was a strong motivation. In addition, the technical challenge of using a VA device could be a barrier to use and we found many relied on the direct phone number to tech support provided by the clinical trial.

Understanding the lifestyle of older adults in a lower-income community helps illuminate these findings. In short, many participants faced challenges during their lives that have made exercise difficult and strong social support critical. For example, prior physical injuries could make exercising painful; prior significant challenges such as losing a home or job could decrease intrinsic motivation to exercise; strong networks of friends and family helped with everything from transportation to understanding technology set-up to health questions. Prior work has shown that older adults from lower-income communities exercise less than those from higher-income communities, and that exercise interventions targeting lower-income older adults are more likely to succeed if they foster social connection in a familiar environment and realistically account for participant lifestyles [89].

We also found that the VA, in combination with social support, provided some participants with additional motivation compared to the paper exercise group. In other words, social support should be a platform to build technology-based health interventions on top of: while social support is a necessary foundation, the utility of the VA technology (e.g., voice commands and adaptable exercise difficulty) is not superfluous. Particularly when designing exercise interventions for lower-income communities, we suggest having the user designate a care partner and providing hands-on professional support to build habits and answer technical questions. Future work could research how long a period of assistance is needed (e.g., perhaps a month of intense assistance could benefit the user in the long-term). Alternatively, an app could be used by a workout class to supplement (but not replace) in-person workouts at home and allow class members to connect socially and encourage each other. The social support must be relatable to the specific target community. To mitigate technological as well as financial barriers, we also suggest that the positive aspects of the exercise app might be achieved using the VA on a mobile phone or smart TV, as opposed to a separate device.

2.6.3 Limitations

Future work could expand the scope of this study. Recruiting participants was extremely challenging and thus we had a participant pool with limited participant numbers and mostly women. Future work could investigate why women seem more interested in this type of study than men. Given the qualitative nature of the study and our goal to highlight issues raised by people in a minoritized community—not to make claims about an entire demographic group—these aspects of our participants' demographics do not impact the validity of our findings but would be worth expanding on. Future studies could explore the differences in experiences of older adults from a variety of cultural backgrounds, geographies, and financial statuses.

2.7 Conclusion

In this paper, we investigate how Black older adults from largely lower-income urban communities perceive and use voice assistants (VAs) for exercise. To this end, we collaborated with geriatric medical researchers who are conducting a clinical trial testing a voice assistant-based exercise app as compared to a paper packet with the same exercises in this demographic group. Two qualitative studies were conducted with clinical trial participants before and after the trial, with a focus on understanding 1) older adults' perception of VAs, and 2) factors motivating older adults to exercise and how these varied between the VA and paper trial groups, respectively. Through the first focus group study conducted before the clinical trial started, we identified diverse ways older adults perceived a VA, which both positively and negatively shaped their willingness to use a VA for exercise. In the second interview study conducted after the clinical trial, we found that while the VA motivated some older adults to exercise, other factors that facilitate exercise like intrinsic motivation and social support were also critical. Informed by our findings, we further discuss how the VA's design could be a facilitator or inhibitor of Black, lower-income older adults' exercise routine.

Focusing on historically marginalized communities, our paper contributes to research on VAs in healthcare for older adults by tackling preventative health maintenance specifically, through a long-term comparative study. Our paper suggests future VAs and their interaction paradigm should be designed to be more culturally contextual, and capable of promoting existing social relationships.

CHAPTER 3

GENERATIVE AI USES AND RISKS FOR KNOWLEDGE WORKERS IN A SCIENCE ORGANIZATION

3.1 Introduction

In this chapter[1] I focus on a project I led understanding generative AI use in a science research organization. Generative AI has demonstrated significant potential to enhance scientific discovery by supporting knowledge workers in science organizations and institutions [8, 179, 182]. However, the real-world applications and perceived risks of generative AI use in these organizations are uncertain. If generative AI could make science organizations with science- and operations-focused employees more efficient and speed up time to discovery on topics such as drug development and climate solutions, it would have important consequences for facilitating scientific gains to help society at large.

Prior literature has either looked at generative AI as a tool for scientific research [182, 194, 62, 115, 150, 164] or generative AI in the workplace, with a particular focus on professional knowledge workers [101, 68, 161, 174, 85, 211, 37]. Very little research, however, studies generative AI opportunities for scientists as knowledge workers [123], and prior research has not studied generative AI for both science- and operations-focused workers in a science organization. In addition, prior work has investigated generative AI risks and concerns (e.g., [68, 136, 141, 123]), however, there has not been a focus on concerns that are specific to the novel context of a science organization.

This work adds to the literature by reporting on an investigation of the practical, real-world applications and perceived risks of generative AI use by employees at Argonne National Lab. Argonne employs several thousand people across Science and Operations teams and

1. A version of this chapter has been published at CHI 2025 with co-authors Matthew Dearing and Marshini Chetty [202].

is tasked with basic and applied science and engineering research. We studied generative AI perceptions and uses for employees spanning Science and Operations roles and measured early adopter usage for a recently released internal generative AI interface called *Argo* based on a private instance of OpenAI's GPT-3.5 Turbo large language model (LLM).

Two features of the national lab make it a useful case study for science organizations and non-science organizations interested in using generative AI. First, numerous knowledge work organizations have a similar dichotomy between knowledge specialists (e.g., scientists, lawyers, investors) and operations workers, and in this research I consider that they might have different needs but similar levels of interest with respect to generative AI. Second, the national lab regularly deals with sensitive data, such as classified, national security, or pre-published scientific data, and thus must take privacy and security risks seriously, similar to organizations such as banks and other government institutions.

Our research questions are the following:

RQ1: How are Science and Operations workers at a national lab currently using, and how do they envision using, generative AI to support their work?

RQ2: What risks (including privacy, security, and ethics) exist for using generative AI at a national lab?

To answer these research questions, I led a team in conducting a survey ($N = 66$) and in-depth semi-structured interviews ($N = 22$) with Argonne employees, and also analyzed Argo usage data from the first eight months of deployment (which overlapped with the study time frame).

There were four main findings: (1) Following its initial launch, we found Argo was being used by a growing number of early adopters (more often in Science) and that most survey respondents were familiar with, and experimenting using, generative AI. However, few had made it a consistent part of their work; We identified common generative AI use cases that conceptually binned into either a (2) *copilot* or (3) *workflow agent* modality. *Copilot* refers to a system that works in conjunction with a user on tasks and provides responses in a conversational manner.[2] *Workflow agent* refers to an autonomous or semi-autonomous AI system that can perform complex tasks mostly on its own to support a user's work. Science and Operations participants reported similar copilot-style interaction scenarios: current uses centered around writing verifiable and structured code or text, while envisioned uses focused on extracting insights from large, unstructured text data. Both Science and Operations participants described how generative AI agents could automate some of their workflows, although these workflows differed between the two roles; and (4) In terms of generative AI risks, we found many participants were concerned about reliability/hallucinations, overreliance on generative AI, data privacy and security, the future of academic publishing and citation practices, and to what extent generative AI would impact hiring and jobs at the national lab.

The novel contribution of this research lies in studying the intersection of generative AI in scientific knowledge work and generative AI in a professional organization, specifically:

- Novel usage data for the deployment of an organization-wide generative AI chatbot at a national lab (powered by a private instance of OpenAI's models).
- A collection of current and envisioned generative AI use cases for Science and Operations employees at a science organization binned into either copilot or workflow agent interaction modalities.
- Novel perspectives from Science and Operations employees on generative AI risks for a science organization.
- Design recommendations for organizational use of generative AI copilots and workflow agents.

2. Here we refer to "copilot" as an interaction modality and do not mean the Microsoft Copilot suite of products.

- Recommendations for future HCI research on generative AI applications in science, and other knowledge work, organizations.

3.2 Background

3.2.1 Generative AI in Science

In recent years, scientists have been exploring how deep learning and generative AI might be useful for advancing scientific research [192, 8, 179]. Some have called for building science-specific large language models (LLMs) trained only on science literature [182, 194]. Another area of interest has been in the use of LLMs to generate science-focused programming code, including calculation kernels executed on HPC systems and parallel code snippets to train LLMs [126, 84, 56]. Generative AI is also being leveraged to create synthetic user research data for Human-Computer Interaction (HCI) research [96, 134] and drive automated process workflows for scientific tasks and experimental equipment [164, 115, 62, 150].

Little research exists, however, on how generative AI is impacting the day-to-day aspects of scientific knowledge work. One exception is a study by Morris [123] who interviewed 20 scientists, at a variety of institutions, and identified generative AI opportunities as well as concerns, ranging from literature reviews and data analysis to applications in higher education given many participants' roles as university professors. Our work differs from Morris' by focusing on scientists at an organization that has a science mission (as opposed to university professors and scientists working in the tech industry). Notably, we include the perspective of Operations workers, investigate organization-level trends, and specifically seek perspectives on privacy and security. Another closely related study did not exclusively focus on generative AI. In this work, Crosby et al. [53] use a human-centered methodology to inform the design of a suite of tools for ocean scientists that leverage ML to process image data. Finally, some work has looked at the ability for LLMs to summarize academic papers for scholars [83, 204] or create data visualizations [118]. We expand on these studies by contributing a comprehensive look at how both Science and Operations staff in a science organization might use generative AI in their work and what concerns they have.

3.2.2 Generative AI in Knowledge Work

Researchers have also begun investigating the use of generative AI in knowledge work, a classification of labor involving the production of information-driven products and services as the key economic output [211, 161]. Some professions considered knowledge work include data science, law, marketing, and finance [161, 211]. Given generative AI's ability to process text-based information, there has been growing interest in how this technology will impact the jobs of professional knowledge workers [101, 68, 161, 174, 85, 116, 117, 211, 37].

One thread of research looks to measure the productivity gains by knowledge workers who have access to generative AI tools [3]. Multiple studies have found that generative AI tools increase productivity more for less skilled workers [33, 129], including for workers in consulting [59] and programming [102, 140, 50]. Other research on generative AI applications in professional settings has largely focused on creative knowledge work in journalism and science writing [142, 105], user experience (UX) and industrial design [198], marketing and public relations [190], and software engineering [160, 50]. Another research direction has investigated creative knowledge work tasks (rather than professions) that include writing [81, 114] or music composition [188]. We contribute to the literature on generative AI in knowledge work by focusing on science as a sub-domain that remains understudied.

In addition to a lack of research on scientific knowledge work, there is also minimal literature in the HCI and Computer Supported Cooperative Work and Social Computing (CSCW) communities studying generative AI at the organizational level. This oversight is significant because there can be networked impacts to organizations (both positive and negative) that might not be felt by individual users. For example, Cortiñas-Lorenzo et al. [48] think critically about how enterprise knowledge systems accessed via AI have the ability to shape and distort how workers view themselves and others, which in turn changes worker behavior. One scenario they describe is a system that allows workers to search for "experts" on a topic in their organization. By labeling some employees experts and others not the system might inadvertently inform who gets promoted or increase workload on certain staff, making it imperative organizations make these systems transparent to workers. We contribute an in-depth case study of a single organization's use of generative AI for knowledge work, considering the interplay between roles (Science and Operations) within the organization as well as risks at the organizational level.

3.2.3 Generative AI Concerns in the Workplace

Alongside the search for generative AI applications in the workplace, researchers have uncovered risks, harms, and concerns relevant for knowledge workers. One of the most prevalent is the fact that generative AI systems "hallucinate" false information with suggestive confidence that it is true [110, 68]. These hallucinations, as well as a lack of citations to original sources, can be particularly concerning for scientists [123]. There is also concern about overreliance, or users placing too much trust that a system is working and not checking outputs thoroughly enough [136, 68]. Overreliance is a concern that predates the development of generative AI extending to early automation systems, such as automated landing systems for airplanes [169]. Thus, researchers are considering how to design for the appropriate level of trust users *should* put in a system [205].

Privacy and security as well as copyright and plagiarism are also work-related generative AI concerns. Some research has found that programmers who used a generative AI coding assistant were more likely to write insecure code [141, 137]. Additionally, significant controversy surrounds the use of writing [156], music [42], and visual art [90, 172] to train generative AI, due to the lack of compensation for artists and authors as well as the ability for the models to regurgitate some training data verbatim triggering copyright laws.

Another important issue drawing commentary from both researchers and the public is the extent to which generative AI will impact workers and industries, particularly risks to knowledge workers' jobs [211, 116, 85]. For instance, Woodruff et al. [211] interviewed knowledge workers in several industries (their sample did not include scientists) about how they thought generative AI might change their field. They found that while participants did not foresee major changes to their industries, there was concern for the rise of deskilling, dehumanization, disconnection, and disinformation. Other research found workers can feel inferior or unaccomplished if generative AI models can largely do their tasks for them [110]. While the spectre of fully autonomous systems has loomed for many industries going back to the Industrial Revolution, there have always remained some aspects of a job that machines cannot do [86]. In this work, we expand on some of these risks and concerns while highlighting how they surface in unique ways both in a science context as well in an organizational context.

3.3 Methods

3.3.1 Research Context

In January 2024, the IT group at Argonne National Lab broadly deployed a generative AI chatbot named Argo, powered by a private instance of OpenAI's GPT-3.5 Turbo,[3] for use by all members of the Argonne National Lab community. Argo was designed for exclusive internal lab use so it does not save query and LLM response data and it does not share such information with OpenAI or other third-party services. Employees could use Argo as they would a service like ChatGPT: it included a browser-based interface that a user could type a prompt into and get a reply. Argo could also be accessed by employees via an API. Argo was intended to be used only for work purposes and required a lab login and VPN to access

3. Argo was upgraded with additional features and access to GPT-4 Turbo after data collection for this study was complete.

if not on-site at the lab. Beyond providing a secure generative AI assistant for employees, Argo was not released with a specific purpose. Employees could find out about Argo through official announcements, emails, or events targeted at raising awareness.

Argonne National Lab is organized into Science and Operations divisions. The goal of Science divisions is to produce scientific research, while the goal of the Operations divisions is to keep the organization running smoothly. Science workers may be operating instruments, running experiments, managing data analysis, writing grants, and publishing papers. Operations workers may be producing science communication, working in administrative roles, ensuring lab safety (e.g. radiation exposure, cybersecurity), and building software and hardware infrastructure. Importantly, while the lab is divided into distinct divisions there is overlap in tasks between them. For example, both divisions employ technical workers such as software engineers. In addition, both divisions require some workers with a scientific background. In this work, we distinguish between the two divisions to tease out if and when the need for generative AI differs between Science and Operations roles since this is a common division of labor in knowledge work organizations.

We conducted a survey (April–June 2024) and interviews (April–July 2024) during the first eight months of Argo's deployment to purposefully capture generative AI perceptions and use levels during this early stage. Our research was approved by the university Institutional Review Board (IRB). Since it is against the law to monetarily compensate federal employees in the course of their work, we did not provide gift cards to participants, but made it clear the research would be used by the lab to improve their systems.

Table 3.1: Survey Participant Demographics

Demographic	**Response Option**	**# Participants (N=66)**	**Percentage**
Division	Science	32	48%
	Operations	31	47%
	Prefer not to answer	3	5%
Role	Scientist	15	23%
	Software Engineer, Data Scientist, IT	15	23%
	Operations Manager	9	14%
	Engineer (hardware, facilities, etc.)	7	10%
	Cybersecurity, Safety	6	9%
	Scientist Manager	5	7%
	Administrative, Communications	4	6%
	Prefer not to answer	5	8%
Years at lab	<1	3	4%
	1-4	17	26%
	5-9	13	19%
	10-14	9	14%
	15-19	9	14%
	20+	9	14%
	Prefer not to answer	6	9%
Age	Under 24	1	1.5%
	25-34	7	11%
	35-44	20	30%
	45-54	14	21%
	55-64	13	20%
	65-74	1	1.5%
	75+	0	0%
	Prefer not to answer	10	15%
Gender	Female	12	18%
	Male	44	67%
	Non-binary	0	0%
	Prefer not to answer	10	15%
Race/Ethnicity	American Indian or Alaska Native	1	1.5%
	Asian	5	7%
	Black or African American	1	1.5%
	Hispanic, Latino, or Spanish Origin	0	0%
	Middle Eastern or North African	0	0%
	Native Hawaiian or Pacific Islander	0	0%
	White	46	70%
	Prefer not to answer	13	20%
Education	Less than high school degree	0	0%
	High school degree	0	0%
	Associate's / Some college degree	6	9%
	Bachelor's degree	13	20%
	Master's degree	20	30%
	Doctoral degree	21	32%
	Prefer not to answer	6	9%

Table 3.2: Interview Participant Demographics

ID	Division	Role	Years at lab	Age	Gender	Race/Ethnicity	Education
P1	Operations	Cybersecurity, Safety	5-9	35-44	M	White	Master's degree
P2	Operations	Engineer	20+	55-64	F	White	Master's degree
P3	Operations	IT	10-14	35-44	M	White	Bachelor's degree
P4	Operations	Cybersecurity, Safety	1-4	35-44	M	White	Bachelor's degree
P5	Operations	Engineer (Facilities)	1-4	35-44	M	White	Bachelor's degree
P6	Operations	IT	1-4	55-64	M	White	Associate's / Some college degree
P7	Operations	Cybersecurity, Safety	5-9	25-34	M	White	Associate's / Some college degree
P8	Operations	Operations Manager	15-19	35-44	F	White	Master's degree
P9	Operations	Cybersecurity, Safety	15-19	35-44	M	White	Master's degree
P10	Operations	Cybersecurity, Safety	1-4	<24	F	White	Master's degree
P11	Science	Scientist Manager	1-4	35-44	M	White	Doctoral degree
P12	Science	Scientist	1-4	25-34	F	Asian	Doctoral degree
P13	Science	Scientist	<1	25-34	M	White	Doctoral degree
P14	Science	Data scientist	N/A	N/A	N/A	N/A	N/A
P15	Science	Scientist Manager	15-19	N/A	M	N/A	Doctoral degree
P16	Science	Scientist	10-14	35-44	M	White	Doctoral degree
P17	Science	Scientist Manager	10-14	55-64	M	White	Doctoral degree
P18	Operations	Operations Manager	1-4	35-44	F	White	Master's degree
P19	Science	Scientist	5-9	35-44	M	White	Doctoral degree
P20	Science	Software Engineer	15-19	35-44	M	White	Bachelor's degree
P21	Science	Scientist Manager	20+	45-54	M	White	Doctoral degree
P22	Science	Scientist	20+	N/A	M	N/A	Doctoral degree

Table 3.3: Themes* and Unique Survey Respondent Counts

Code	Code Description	Survey Respondents (%**)
Current/envisioned uses for LLMs		
Writing structured text/code	Use cases related to writing structured text/code	13 (20%)
Query unstructured data	Use cases related to querying unstructured data sets	13 (20%)
Workflow automation	Use cases related to workflow automation	18 (27%)
Other	Other use cases	5 (7%)
Issues or ethics concerns at work		
Reliability	Concerns about whether LLMs are reliable, trustworthy, etc.	29 (44%)
Overreliance	Concerns about people relying too much on LLMs	14 (21%)
Academic publishing	Concerns about academic and science communication publishing	13 (20%)
AI taking jobs	Concerns about the impact of LLMs on human jobs	3 (5%)
Social concerns	Broader social concerns that participants do not tie directly to their work	9 (14%)
No ethics concerns at work	Participant did not have ethics concerns about LLMs at work	22 (33%)
Privacy/security concerns at work		
Concerned	All concerns related to privacy and security	28 (42%)
Not concerned if using organization tools	Participant not concerned about privacy/security as long as they were using LLMs officially designated secure by their organization	12 (18%)
Concerned but prefer commercial LLMs	Participant felt organization-designated LLMs were safer but preferred the features available in commercial LLMs	3 (5%)
No privacy/security concerns at work	Participant did not have privacy/security concerns about LLMs at work	22 (33%)

- *Bolded codes correspond to the interview codebook codes that we analyzed for this work. Sub-codes correspond to themes used to code the survey short responses.
- **Because not every survey response included each theme, and some included multiple themes, percentages do not sum to 100%.

3.3.2 Data Collection

Survey

We first wanted to broadly capture national lab employee perceptions of generative AI uses and concerns in a survey. The survey, created in Qualtrics, included questions such as *How familiar are you with large language models (LLMs)?*, *How often do you use LLMs as part of your work?*, and *Please describe ethical concerns you have about using LLMs in your work, if any.* We also asked *How often do you use LLMs for the following work tasks?* and provided a list of 15 tasks such as *writing code*, *feedback on experimental design*, and *editing human-written text.* This list of tasks was drawn from a study of 20 scientists about their use of generative AI [123]. We also asked demographics questions. At the end of the survey, we gave participants the option to share their email if they would be willing to sign up for an interview. See the full survey in the Appendix of the paper [202]. We pilot-tested the survey with a small set of HCI experts in Computer Science and refined wording and ordering of the questions. After finalizing the changes, we deployed the survey to Argonne National Lab employees via mailing lists and at presentations given by our collaborators at the lab.

We received a total of 80 survey responses. We filtered out surveys that were incomplete, with the exception of three that only lacked demographic information (given that a valid option was "Prefer not to answer" we selected this option for these responses) or short answer responses, but had all the multiple choice questions completed. We also removed one response that was disingenuous based on the answers. After filtering, we were left with 66 responses to analyze.

Interviews

As survey responses were completed, we contacted respondents who had provided email addresses for follow-up interviews. In total, we contacted 40 respondents, 22 of whom agreed to participate in semi-structured interviews. Each interview lasted 30 minutes and was conducted over Zoom and recorded. The interview protocol was designed to elicit more depth on generative AI applications, risks, and concerns than the survey. We asked participants to recall current or envisioned scenarios for generative AI, and also prompted them based on their survey responses. We then went in-depth on each scenario, asking questions such as *To what extent have LLMs been helpful for this task?* and *When using an LLM to do this task doesn't work, what goes wrong?* In addition, we dug into participant views on privacy, security, and ethics concerns.

Argo Usage Statistics

Between January and August 2024, we collected high-level metadata during the regular use of Argo including authenticated user name, time of use, LLM selected and related configuration options, and the size of user queries and responses as measured by token counts.[4] The text-based content of these queries and responses are not automatically stored by any external or internal database system. Only an Argo user may optionally save their LLM-based conversations to their local machines before disconnecting from their session with the service through a copy-to-clipboard mechanism or a transcript download feature. We roll-up the tracked user name to their organizational division, which is then categorized as either being within the Science or Operations workforce group.

3.3.3 Participants

Table 3.1 shows the demographic information for the survey participants. We received approximately equal numbers of Science and Operations employee responses. Most respondents were scientists or engineers. Respondents had been at the lab for a range of years, includ-

4. A *token* is unit of data processed by an LLM, such as a word, a part of a word, or a collection of symbols. A count of the number of tokens representing a block of text is a simple and standard measure of the size of that text, without exposing its content.

ing both new and highly seasoned employees. We also represent a range of ages, primarily between 25 and 64 years old. The gender and race/ethnicity breakdown is representative of the lab population. Most (but not all) respondents had at least a Bachelor's degree, and many had advanced degrees.

Table 3.2 shows demographic information for interview participants. All interview participants also provided a survey response, and the demographic breakdown is similar. It is skewed toward white men, although this is reflective of the overall national lab population. Throughout the work, we denote interview participants IDs with "P" and survey participant IDs with "S."

3.3.4 Data Analysis

Interviews

We had the interviews transcribed using Rev.com under a non-disclosure agreement. We qualitatively coded the interview transcripts and then analyzed them using thematic analysis [162]. First, one researcher read through the transcripts and created an initial codebook based on the interview questions. The team then discussed and refined the codebook and finalized the codes (the codes we analyzed for the final work are bolded in Table 3.3, see the full initial interview codebook in the supplementary materials). One researcher subsequently coded all the transcripts using MAXQDA. Then, the same member of the research team used axial coding to assign sub-codes through an iterative process to the following parent codes: *Current/envisioned use cases for LLMs*, *Privacy/security concerns at work*, and *Issues or ethics concerns at work*. Another member of the research team then did a second round of coding, reviewing all sub-codes labeled in the first round. All points of disagreement were discussed and resolved between the two coders. The team also met regularly after a set of transcripts were coded to ensure the coding process aligned with the goals of the study. After coding, two members of the research team extracted themes and discussed these with the full team (see themes in Table 3.3 as well as the interview codebook found in the supplementary materials) and further refined those that were most related to the research questions.

Survey

We calculated descriptive statistics for the survey based on the multiple choice responses using Python. We qualitatively coded the short answer responses in MAXQDA. We began the coding process after extracting themes from the interview data. Thus we used the same codebook for the survey, as shown in Table 3.3, along with the unique number of survey responses for each code.

3.4 Findings

Our findings represent the perspectives of generative AI early adopters at the national lab. In Section 3.4.1, we describe initial generative AI usage at the lab, showing that less than 10% of all lab employees used Argo each month in the study period but that there is an upward trend in use. This is a lower bound for generative AI usage more broadly because in our surveys, more employees reported trying other commercial LLMs such as ChatGPT.

Drawing from both the survey short responses and the interviews, we identified common *current* and *envisioned* use cases for generative AI and conceptually split these findings into two categories: *copilot* (Section 3.4.2) and *workflow agent* (Section 3.4.3). We review similarities and differences between Science and Operations workers with respect to the kinds work they want to accomplish in each category. In Section 3.4.4, we highlight the most common risks and concerns for using generative AI in the national lab that surfaced in both the survey short answers and interviews.

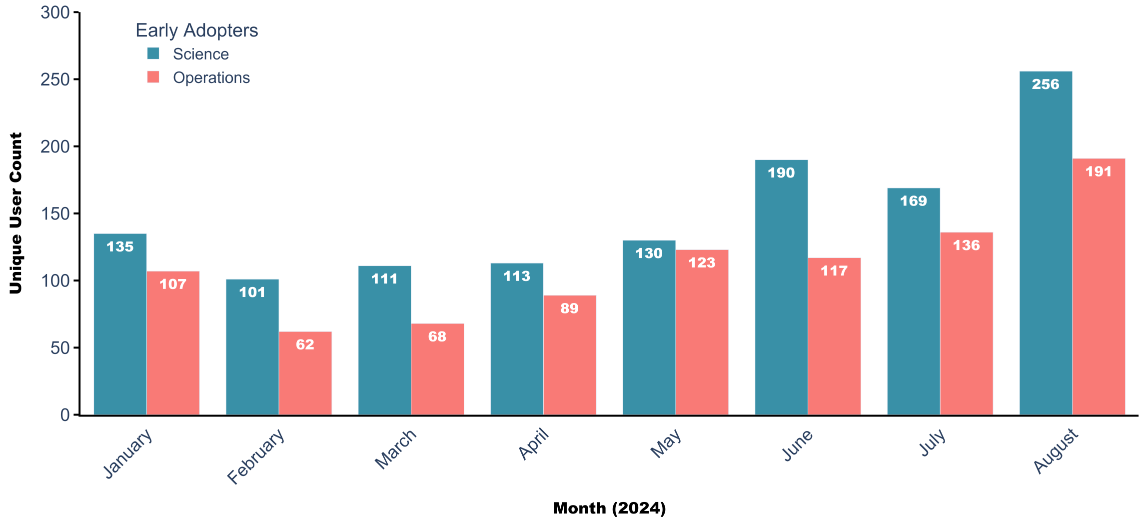


Figure 3.1: Argo usage metrics by Operations and Science unique users for each month since initial deployment. Monthly usage is less than 10% of all lab employees. *Note: this plot does not capture employee use of external LLMs and is based on auto-collected telemetry data.*

3.4.1 Generative AI Usage and Familiarity

We found that early adopters are familiar with generative AI and using it experimentally in their work from our survey data and Argo usage data. Figure 3.1 shows an upward trend in usage[5] during the early months of launching Argo. While the use period seen in Figure 3.1 is brief for designating trends, we observed a general increase in generative AI use across Science and Operations during the study period. Specifically, unique users of both groups, excluding the initial launch month, increased an average of 19.2% each month, broadly ranging from approximately no change to as high as nearly 47%. Across this same period, monthly unique users in Science and Operations increased 158% and 200%, respectively (or 174.2% over all users between the second and final months of the reported statistics). While there were 26% more unique monthly Science users on average compared to Operations users, we observe an *average monthly increase* in the latter group at 21.2% compared to 19.3% for Science users.

5. The usage data is for all users, which may include survey respondents but does not correspond directly to them.

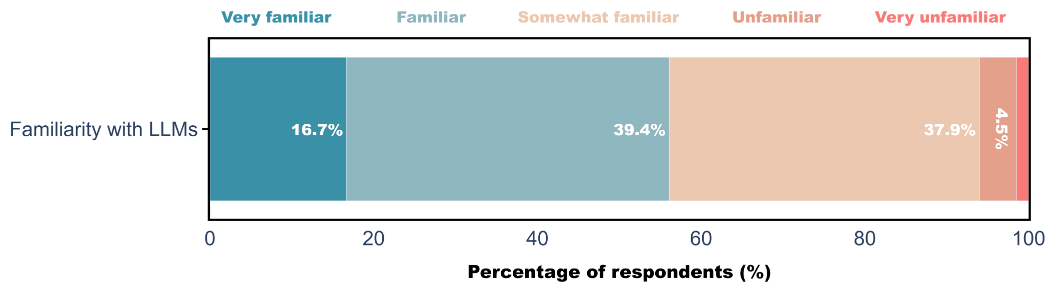


Figure 3.2: Responses to the survey question: *How familiar are you with large language models (LLMs) such as ChatGPT, Argo, etc.?*

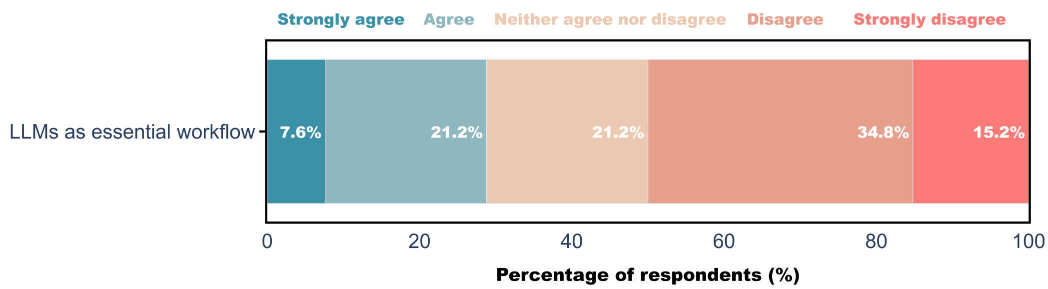


Figure 3.3: Responses to the survey question: *To what extent would you agree/disagree with the following statement: LLMs have become an essential part of my workflow.*

Overall, we see a suggestive early trend that the availability of a secure internal generative AI solution is of significant interest to Science and Operations users at a national lab, and this data provides a useful baseline for future studies.

Most of our survey respondents, and all our interview participants, had some degree of familiarity or experimentation with generative AI. Based on the survey short responses and interviews, participants also had a sophisticated understanding of how generative AI models are trained. In our surveys, we asked employees about their familiarity with, and use of, LLMs more broadly. Figure 3.2 shows that 94% of participants had some level of familiarity with generative AI tools, although only 16.7% felt they were *very familiar*. Similarly, despite high levels of familiarity, only 28.8% of survey respondents said they *agreed* or *strongly agreed* that LLMs had become essential to their workflows (Figure 3.3). The majority of survey

Table 3.4: Responses to the survey question: *Which LLMs have you used before?*

LLM	Respondents (%)
ChatGPT (OpenAI)	54 (82%)
Argo	44 (67%)
Gemini (Google)	17 (26%)
LLaMA (Meta)	11 (17%)
Claude (Anthropic)	6 (9%)
Other	7 (11%)
None	7 (11%)

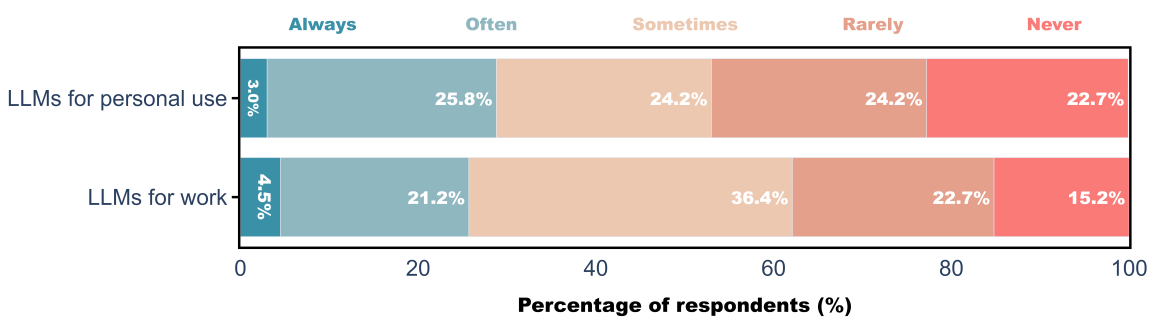


Figure 3.4: Responses to the survey questions: *How often do you use LLMs as part of your work?* and *How often do you use LLMs for personal use?*

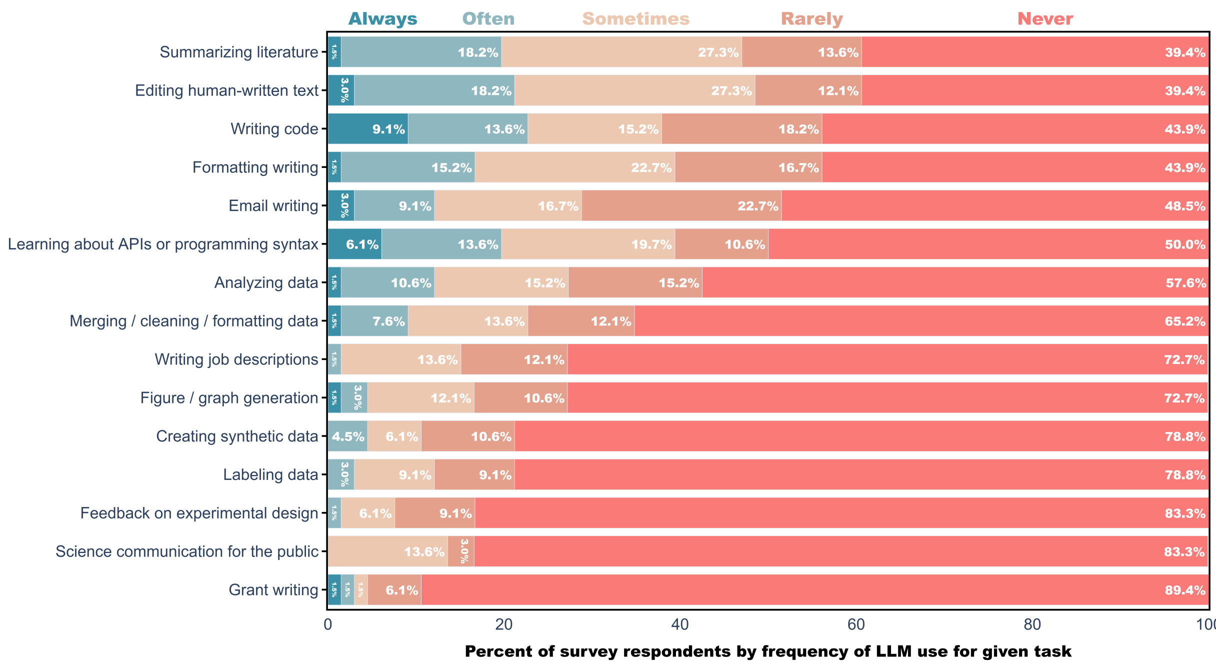


Figure 3.5: Responses to the survey question: *How often do you use LLMs for the following Argonne National Lab-related work tasks?*

respondents had primarily tried ChatGPT (82%) and Argo (67%) as shown in Table 3.4. Our survey respondents also did not report a large difference between how much they used LLMs for work and personal tasks as shown in Figure 3.4.

Figure 3.5 shows how often survey respondents used LLMs for a series of tasks in the context of their work. Nearly 60% of respondents had at least tried using LLMs for *summarizing literature*, *editing human-written text*, *writing code*, and *formatting writing*. While the fewest total respondents (11%) had tried using LLMs for *grant writing*, we note that one person responded *always*, suggesting that it might be that respondents want to use LLMs for these tasks but have not figured out how to given their relative complexity. Examples of personal generative AI use that came up un-prompted in the interviews included scripting narratives as part of the game Dungeons and Dragons (P5, P13, P19) and creating furniture arrangements for a floor plan (P12). This data suggests early adopters are in an experimentation phase.

Table 3.5: Generative AI Use Case Examples in Science vs Operations Roles

Employee Copilot		
Use Case	*Science Use Case Examples*	*Operations Use Case Examples*
Writing structured code/text	Writing academic paper introductions, grants, reports, emails, and code	Writing reports, emails, and code
Extracting insights from large unstructured text data	Extracting insights from scientific literature; querying lab rules and regulations	Extracting insights from public data sources such as specified websites; summarizing team data such as surveys; querying lab rules and regulations
Workflow Agent		
Use Case	*Science Use Case Examples*	*Operations Use Case Examples*
Initial steps towards automated workflow agents	Operating and extracting data from scientific instruments; automating data analysis pipelines	Automating instrument safety checks
Fully automated workflows	"AI scientist" that can generate hypotheses, run simulations, and revise hypotheses based on output with minimal human input	Complex project management, such as prioritizing tasks, creating Gantt charts based on team updates, and long-term planning

3.4.2 Science and Operations Employee Copilot

We categorize the first set of generative AI use cases described by participants in both Science and Operations roles as *copilot*-style interactions. This means they entail conversational interactions between the user and the AI where the user gets real-time responses to questions posed to the AI. We use the conceptual framing of a copilot in order to arrange our findings around features and affordances this copilot would need to be most useful in a science organization. We note that participants themselves rarely used the term copilot, rather we impose it for conceptual organization of the findings.

At the time of data collection, participants said they were most often using LLMs such as ChatGPT to get help writing structured text that they can easily verify is correct, such as emails and reports. Participants envisioned goal, however, was to use a LLM to extract insights from unstructured text data such as scientific literature or survey results. We group Science and Operations employee responses together in this section since we did not find a large difference in use for copilot-style interactions.

Current Uses: Writing Structured Code/Text

Survey and interview participants described numerous examples of structured writing that could be aided by LLMs such as creating emails, reports, grants, and the introduction to academic papers. By structured, we mean types of writing that follow standardized formats. Participants described that they typically already had the content needed for the writing, but they used LLMs to craft the the appropriate tone and format. In the survey, over 50% of respondents said they had tried using LLMs for each of the following related tasks: *editing human-written text*, *writing code*, *formatting writing*, and *email writing* (Figure 3.5). Moreover, 20% of short answer responses included use cases related to writing structured text or code (Table 3.3).

When writing emails, participants most commonly described using LLMs to help with tone given a specific audience. P3, a senior IT employee, summarized the feelings of many participants saying that he would provide the LLM a document or paragraph and ask it to, *"make this more formal, make this less formal, make this friendlier"* to reduce his own emotional labor. Many participants described wanting to sound "professional." For example, P18, an operations manager, said she asks the LLM to *"make this sound better or make this more professional or make this more succinct."* Similarly, P21, a scientist manager, said he used ChatGPT to make emails *"more formal and less scientific."* P16, a scientist, even explained that he could get frustrated but did not want that to come across in his emails so, *"AI writes the email so that I don't have to respond with the emotional outrage I might have."* P12, a scientist, noted that as a non-native English speaker, she found this ability to set the tone with the LLM particularly helpful and several survey respondents mentioned using LLMs for language translation.

In addition to writing emails, participants also described how LLMs were helping them write other kinds of structured text such as internal reports. The national lab, like many workplaces, requires specific types of reporting from employees. P17, a scientist manager, described this saying, he was required to write a specific project document and that the national lab had provided *"a guideline that is like a template"* and he felt that *"the AI can prompt you and make sure that you're giving it all of the paragraphs that it needs and then have it summarize the report for you."* P2, a senior employee in Operations, also found LLMs to be useful in writing internal reports saying, *"Who wants to write these things, right? Who has the time? But... if you take an incident debriefing and ask the system to write a [report] based on that... it did a really nice job."*

Scientists also mentioned grants and research paper introductions as other forms of structured writing LLMs could help format. P11, a scientist, described the effort needed to write grant proposals as "enormous" and that *"anything that saves time in that process is potentially hugely valuable because it lets the researcher then focus that time on the actual research... I* *think there's a potentially big advantage if we could get to the point where there's a useful place for LLMs in that process."* P16, P19, and P22 all thought LLMs could take a research paper draft and summarize the content to draft the introduction to the paper. To reiterate, these were not cases of the LLM generating the research content, but reframing existing content to fit the typical structure of an research paper introduction (in Section 3.4.4, we further explore ethical concerns related to AI authorship).

Envisioned Uses: Extracting Insights from Large Unstructured Text Data

As science-focused knowledge workers, national lab employees must process significant quantities of unstructured textual data, such as scientific literature or organization regulations. At the time of the study, employees were hesitant to trust generative AI with extracting insights due to fears of hallucinations and reliability, which we return to in the Section 3.4.4. If these issues were resolved, however, we found a strong desire among participants to be able to get help from generative AI with managing, organizing, and learning from designated data sources. In the survey, 60% of respondents said they had at least tried *summarizing literature* (Figure 3.5) and 20% of survey free responses mentioned use cases related to querying unstructured text data (Table 3.3).

Multiple interview participants envisioned interacting with unstructured data in a conversational manner. In a representative example, P3, a senior Operations employee responsible for anticipating the software needs of scientists, said *"I think the game-changing aspects would be in interactive conversational simulation modeling."* He continued with an example where a scientist could *"have brainstorming sessions with the system itself, almost replicating the type of thing that happens in colloquia or in focus discussion groups, but making one of those partners an LLM system."* In this section, we cover the types of text data participants were most interested in.

Scientific Literature: Scientists, in particular, wanted a tool to help search and summarize scientific literature. P11, a scientist, made a representative comment: *"And like every other academic in the world, we have to do a lot of lit reviews, and that's hugely time intensive and it's just tedious"* and he envisioned that once LLMs became more accurate they could *"get each paper down to a set of bullet points, that could be a huge time saving."* P22, also a scientist, echoed this perspective and pointed to the increasing amount of scientific literature saying, *"Because nowadays, literature, there's too much literature, too many papers. So that could be very helpful. Maybe it can go out, go to other sources, and then in a particular field, summarize, this week, what happened in this field?"*

Public Data Sources: Lab employees who were not directly involved in writing research papers were also interested in the ability to mine public data sources to extract insights. For example, P4, an Operations employee, wanted to be able to track broader trends online to help with security-related tasks. He described his ideal tool saying that he wanted to be able to create *"a customized library where we can add resources to that library and then basically ping those things [with a LLM]."* P4 described that currently, this kind of public knowledge synthesis for security purposes is done mostly manually. In addition to being able to query data, P7, another security specialist, said he wanted to know how these kinds of public trends related to research being done at the national lab.

Team Data: At the team level, participants wanted the ability to extract key points and notes from meeting transcripts. P10, an IT employee, said she helped process employee requests for approval to use third party tools and that many of these requests were tools that, *"would transcribe the text of the meeting, but it would also create brief notes or overview outlines of what meetings were about and takeaways from that."* P9, a lab safety manager, echoed this desire and explained that he was part of a weekly supervisors meeting and he wanted to be able to store and query the meeting transcript for summaries, context on a topic discussed, action items, and other similar questions. Some teams had surveys that they wanted to analyze, for example, P8, an Operations manager, described a situation where her team collected 313 responses from the lab about a safety incident and they used Argo to help identify themes.

Organization Data: At the organizational level, many participants were eager to be able to more easily search lab-wide rules and regulations, including everything from vacation policies to science lab safety standards. P19, a scientist, had a representative perspective saying, "*We have a lot of procedures and documents and policies and all that stuff that's spread all over the place. I'd love to see [LLMs] be used for that.*" Survey respondent S32 wrote that the types of national lab regulations they wanted to search included: *"policies, directives, executive orders, contract requirements."* Participants also described the need for combining existing lab datasets. P8, an Operations manager, said, *"right now our systems are very siloed and are not integrated very well... So I see... potentially doing some kind of [LLM] analysis there."* S37 also pointed to the capability for LLMs not just to search for lab-wide information, but to surface correlations across documents.

3.4.3 Science and Operations Workflow Agents

We categorize the second set of generative AI use cases described by participants as *workflow agents*. As opposed to a copilot, an agent navigates a complex task autonomously or semi-autonomously and returns the output to the user. In the context of a science organization, we found agents were emerging as a way of driving workflows in both Science and Operations. Scientific workflows included steps such as downloading data from an instrument or database, running multiple data analysis steps, and producing graphs or other visualizations. Operations workflows included tracking if work is progressing on-time, managing procurement processes, and automating common database interactions.

Tasks related to workflow agents that survey respondents had tried included: *analyzing data* (43%); *merging, cleaning, formatting data* (35%); *figure, graph generation* (27%); *creating synthetic data* (21%); and *labeling data* (21%). In the survey free responses, 27%

of answers (equally split between Science and Operations respondents) mentioned workflow automation as a use case for generative AI.

Current Uses: Initial Steps Towards Workflow Agents in Science and Operations

Participants in both Science and Operations described cases where they were testing using LLMs to automate some of their workflow. Participants reported LLMs are already able to automate some of these workflows in a scientific environment to make them more efficient, but generative AI workflow agents are in early stages of development and use.

Science workflows: Multiple participants described how they were beginning to use LLMs to automate their customized scientific workflows. P21, a scientist, described how he is starting to use LLMs to automate getting results from a specialized instrument. His research uses lensless imaging, in other words, rather than take a picture with optics *"you can do lensless imaging where you... just record how is the light or the x-ray scattered from the sample."* In order for the data to be useful, P21 explained that researchers use an algorithm to reconstruct the data to create an image, but this is a complex and compute-intensive process that requires *"experts that really massage the data and tune parameters in order to make this work well."* He began experimenting with using LLMs to help with this reconstruction process. He used the analogy of editing a photo using Photoshop saying, *"Think about Adobe Illustrator or Photoshop, right? Let's assume you've got a picture of someone and the background isn't great... So yes, you can play with Photoshop until you get that image about right, but with these large language models... where you can interact with the chatbot and say well can you sharpen the image or change the tone or whatever it is that you want."* While the technology is not fully functional yet, he sees a future where researchers can similarly ask for a scientific "image" to be sharpened without manually fine-tuning every parameter.

P14 is a data scientist on the same Science team as P21 where the research used lensless imaging; he explained that he helped researchers use computational imaging techniques and that he was building an automated data analysis workflow using LLMs. He said that part of his data analysis process involved setting over 20 parameters in a script, and even though he wrote the script he still forgot the names of some of the variables, so he set up an LLM to *"convert natural language [parameter names] to Python code or MetaLab code, which directly feeds to the computer for actual data analysis."* P14 continued that in addition to generating code, he provided the LLM with knowledge of the analysis techniques he uses, since the LLM's understanding of the techniques on its own was not detailed enough. He said that the knowledge is *"essentially a bunch of text files... a summary of my past experience with different techniques.... whenever I have new data coming in, I want to use the workflow for the new data. And as I process the new data, I pay more attention on my thought process. So in the end I summarize them into the knowledge file."* P14 and a colleague were the main users of his automated workflow, however, he said it is *"actually more useful to inexperienced users or people who just started to learn the technique."* While P14 was optimistic about the ability for LLMs to automate large swaths of his workflow, he said they were still limited to "narrow environments."

Operations workflows: At this point in time, only the more technically-oriented Operations employees had actually tested creating agent-driven workflows. P1, a safety expert, said he was automating workflows for his group, including building *"an application for performing instrument checks in the morning."* He described that since *"a lot of the [instrument checking] software involves these really old scripting-based inputs from the 1960s, no one has a helpful GUI anymore."* So, he designed a workflow agent where he could *"select, 'I want this [chemical element], I want these parameters,' and then just click run"* and the LLM automatically generated a Python script that controlled the old software programs. P1 explained that he did not have any formal training in software engineering but in addition to using a LLM to drive the automation, he also used it to design the Python script: *"I would* *never have been able to [write the code] without significant time investment, and the fact that I could produce a working app in a couple of days was impressive to me."* In addition, P1 described a database his team maintained of relevant chemical elements. He explained how LLMs could be useful for automating searching this database as part of workflows so that his team did not need to export to Excel and do their calculations manually. He wanted to be able to ask the database using an LLM, *"Hey, there was an incident or a fire in this lab. What's the total amount of activity that was in this lab?"* While P1's use cases are specific to his group, the broader ideas he described apply to numerous Operations roles such as: automating old software tools, writing an automation software program without coding skills, and simplifying database searches.

Envisioned Uses: More Fully Automated Workflow Agents

Participants in Science and Operations envisioned uses for generative AI that were more fully automated extensions of their current uses.

Science workflows: Researchers at the national lab have been working towards science-specific large language models that could be integrated into an "AI scientist" tool. As P3 and P20 described, an AI scientist is an advanced workflow agent able to produce scientific simulations that generate new hypotheses that get tested immediately in a feedback loop with minimal human intervention. P20, a software engineer on a Science team, explained his understanding of the initiative: *"[there is] a humongous search space in different science domains, whether it be material science or biology or something like that, and trying to give guidance to researchers so that they don't waste time or it's not just like throwing darts at a dart board."*

Operations workflows: Multiple participants in Operations roles envisioned how LLMs could more fully automate their workflows. S60 provided a representative comment that they wanted LLMs to, *"automate more things—auto organize email, create auto-responses* *to emails, help organize my task list(s) by prioritizing and setting dependencies (can't start task B until task A is complete)."* P17 is a senior manager in a Science division in charge of upgrading a major experiment, however, his day-to-day work aligned more with operations than research tasks.[6] During the multi-year project to upgrade the experiment, he said there were safety talks every day that *"would outline which teams were in which areas on which days."* He envisioned how a workflow agent could use the slides and content from these talks to *"summarize the tasks for all of the teams and create a Gantt chart"* that he could then compare against the timeline that was originally planned in order to *"tell me what took longer and tell me what was faster than what was planned."* Overall, participants envisioned that workflow agents could automate operations tasks related to communication and project management.

3.4.4 Risks and Concerns

In this section we cover the most common risks, concerns, and barriers to using generative AI that participants mentioned. We focus specifically on these topics in the context of participants' work in a science organization. Some interview participants and 14% of survey respondents mentioned broader social concerns—such unauthorized web scraping for training data and bias in datasets—but since they did not tie these back to implications for their own use of generative AI in their work we do not include them in the discussion.

Reliability

Our findings show that the most significant barrier to adoption in a science organization is generative AI's lack of reliability and tendency to hallucinate, as well as the fact that it does not cite source material. Many interview participants and 44% of survey respondents

6. We note that in an organization like the national lab, Science and Operations roles can be interrelated, which is part of why studying both sides is important.

mentioned reliability concerns. One representative survey respondent (S19) said that LLMs were, "*not 'smart' enough to properly search, cite, summarize literature. Issues include incorrect information and making up fake citations.* " P4, a lab safety employee, summarized this viewpoint saying, *"I need to be able to tie [an LLM output] back to an authoritative source."* Citing sources was particularly important for scientists, who noted that current LLMs were not reliable enough for technical information unless they were used by an expert who could spot the falsehoods.

Overreliance

Many interview participants and 21% of survey respondents were concerned that the introduction of LLM tools would lead to overreliance or inappropriate use cases. Representative survey responses included, *"I'm concerned that people who don't know enough to realize the LLM isn't always correct will use the information as if it's true."* (S39) and "*[LLMs are] a tool and over-confidence in their inherent ability will introduce a bias in the that LLM is 'right' or 'correct.'*" (S37). Multiple participants said they were particularly concerned about the use of generative AI in the hiring process, with S29 giving the example, "*for instance feeding resumes or communications to the LLMs and having them making a [hiring] determination on a more qualified candidate.*" Participants who felt they understood how generative AI works were not concerned about overreliance issues for themselves, but for others who they thought might not understand. This was the case for both Science and Operations participants. For example, S18 said, "*I've seen very smart researchers try to use it as a search engine to verify information,*" which is problematic because LLMs can hallucinate false information.

Privacy and security for unpublished, classified, and proprietary data

All of our interview participants and 42% of survey respondents had some degree of privacy and security concerns with LLMs in their work. P12, a scientist, broadly summed up this perspective saying, *"I do not have fear for the technology itself, just to not know who will use it and who will have access to the data. That seems what I am worried about, especially about the datasets."* While participants agreed there were concerns about feeding data into commercial LLMs, they differed on exactly where to draw the line, particularly with unpublished academic research.

Lab employees handle both classified data as well as sensitive data such as personally identifiable information (PII), and participants did not feel it was safe to put these kinds of data into commercial LLMs. For example, P13, a scientist, explained *"Parts of the project that I work on are controlled information, so obviously I couldn't be posting anything like that there."* P11, a scientist manager, felt the same way about public and commercial LLMs, but thought the national lab's private LLM instance *"might even be okay for most types of controlled and classified information."* However, he noted more work would be needed to ensure the safety of classified information, *"Of course, there are other use cases where it's more sensitive than that. That would be a whole different conversation, because I have to be running in a separate [secured digital] environment, things like that if we're actually dealing with classified information."*

Participants disagreed on the extent to which their own unpublished research should be shared with commercial LLMs. At one end of the spectrum, P13, a scientist, did not want to share anything back to companies like OpenAI, saying, *"I would never post, for example, portions of my code and then ask why does this not work? Or something like that."* In contrast, many participants did feel comfortable sharing code snippets and in fact this was a common use case. The national lab has policies regarding "export control," or controlling how research is published to the public. Several participants brought up that putting an unpublished paper into a commercial LLM might constitute publishing it publicly and thus fall under export control policies, but overall there was uncertainty about this issue. On the other hand, some scientists said that without context their results were useless and that they were planning to publish them soon anyway so it was not an issue. For example, P16, a scientist, said he was not concerned about putting his unpublished research into an LLM because *"if it's proprietary, I don't use [commercial LLMs]. And if it's not proprietary, I'm about to publish it so that everybody can read it."*

For many participants, using Argo eased privacy and security concerns because it used a private instance of OpenAI's models. For example, P5, a facilities engineer, voiced a common feeling that the lab could be trusted more than companies: *"In a lot of ways I kind of trust [Argo] more than Claude or... ChatGPT. I trust [it] to keep my personal information more secure than I trust a private company."* Similarly, P6, who works in IT, said he did not have security concerns with Argo because, *"It's contained within our firewall, so I don't have any concerns with that. I think if we would use another LLM that I would have privacy concerns, but not with [our instance] in particular."* S19 said, *"My work involves sensitive information and novel experiments, and so I am apprehensive to provide too much detail to [non-Argonne National Lab] LLMs."* At this time, the private instance also does not store chat history. P7, a lab safety employee, said this was good for privacy and security, explaining, *"I think the best thing is that [chat history] doesn't get stored anywhere because... even in [the national lab] there's going to be probably [several thousand] people who could access it. So with every different use case in science, admin, and stuff like that. So I think making it so if someone puts something in that maybe shouldn't be, the fact that it won't be there is good from a more privacy focused [angle]."*

However, this strategy of promoting Argo could backfire if employees prefer other more advanced LLM models that are available. At the time of the study, OpenAI had released GPT-4 but only GPT-3.5 was available through the national lab's system. P9, a lab safety manager, explained this saying, *"I think the capability of that if we're hosting it here on site... is just so much more limited to what's there. The natural attractiveness of going outside of [the national lab's] intranet is just obvious based on capability."* P21, a scientist manager, echoed this saying he preferred using ChatGPT: *"I've been using chatGPT as opposed to the [Argo] thing. So far, I have not used it with any data where we have any concerns with regards to security, privacy, export control, or any of these topics."* While participants who were using public models said they were being careful, from an organizational perspective this could be a security threat.

Academic Publishing in the Era of LLMs

Many scientists and 20% of survey respondents were concerned about researchers using AI to write academic papers. However, they did not agree on exactly where to draw the line on acceptable vs unacceptable use. On the one hand, everyone agreed that generating fake research was unacceptable. For instance, P21, a scientist manager, said, *"Some individuals... just create fraud... AI and machine learning opens that up to a whole new level because you may be able to really create papers at a push of a button."* Similarly, P19, a scientist, said, *"I think there is also an ethical concern of people using something like a large language model to write a paper and then publishing it and claiming it as their own work. So there's a plagiarism aspect there."* Another scientist, P16, said he had been asked to review some AI-written work for a journal and had suggested it be rejected.

Part of the issue, participants suggested, is that if something in the paper turns out to be incorrect it is not clear who to blame. Also, if the LLM suggests something useful, it is not clear how to cite it. For example, P19, a scientist, said, *"If I write a paper and I submit it and it turns out that I said some stuff that was wrong, well at least there's a person we can have that conversation with. If it comes from a chatbot, then what do you do with it?"*

Most participants felt some level of paper editing with a LLM was acceptable, but the line was fuzzy. P13, a scientist, outlined this tension felt by many saying he was "troubled" by *"the idea of generating content that you pass off as your own"* while at the same time he felt that if ChatGPT was just used to help minimally refine a researcher's original idea then it was overkill to *"put a disclaimer at the front of your work that says this was partially or fully AI assisted."* Overall, participants saw a role for LLMs in writing research papers, but we found that it was still an open question as to exactly how much assistance would be viewed as acceptable.

Impact of Generative AI on Jobs

Participant opinions were mixed on the impact generative AI would have on jobs. Many interview participants brought up the topic, and we asked managers about it directly. While it was only mentioned by 5% of survey respondents, some felt strongly it was important. For the most part, participants saw generative AI as a tool that would be helpful to but not replace scientists, but there was some concern for jobs that required less scientific expertise, such as Communications or IT roles. Many managers believed the skill sets they hired for would change but not the total number of workers.

Participants in technical or specialized roles in Science and Operations tended to view generative AI as a tool that could speed up their infinite to-do lists, but that would not replace them. For example, P17, a scientist manager, said *"[generative AI] will just be another tool like FEA [finite element method] analysis, right? FEA analysis is faster than doing it by hand and more thorough [and LLMs will] be like that."* Another scientist manager, P21, similarly said, *"It's a new technology that will have a very significant impact on the workforce, but essentially it's like a tool."* Likewise, P16, a scientist, said *"It's giving me new tools to use."* Most technical workers were not concerned about losing their jobs due to generative AI. P9, a safety manager with a technical background, summarized this saying the efficiency gained from using LLMs, *"opens up time to do other things that wouldn't have gotten done"* and that he was *"completely comfortable with the idea that it's better at [some tasks]... I have no concerns of that taking my job away."* P16, a scientist, said that having a LLM be able to do programming tasks was like *"having a free junior scientist"* that his group would not have had the money to hire a human for in the first place. However, he said much of his job required being on-site at a scientific experiment *"to make sure that a laser hits the right spot"* and he felt *"AI is never going to be able to do that well, at least not for a very long time."* P18, who works in Operations, said, *"Nothing [in our work] is general. We're not doing general, we're solving a problem based on a specific set of concerns"* and therefore that she was not concerned *"about it taking over people's creativity of their work application, yet at least."* Only one technical participant, P14, thought AI might take his job as a data scientist one day but seemed thrilled by the possibility rather than afraid saying, *"My dream is to develop a fully automated workflow that kind of replaces me... Honestly, I think a lot of my job can be replaced by LLM given enough time and training."*

On the other hand, numerous participants in both Science and Operations were more concerned that AI might take over less technical or specialized jobs. For example, P6 works in IT and his job was largely managing existing automation and fixing issues as they arise: *"I will update correct data in the database tables, troubleshoot issues with some applications for production support, upgrade small commercial off-the-shelf products."* Given that it is conceivable LLMs will be able to automate some of these tasks, he was concerned, saying *"I think that somewhere in the near distant future they're going to put people out of work... you're going to have somebody that is in my [older] age group that's maybe stubborn and they don't want to learn technology."* P16, a scientist, thought AI might take jobs that are largely reading or writing based saying if *"your job is to attend meetings and write emails, AI could replace 90% of what you do."* P19 was concerned about AI taking jobs but not research jobs that required "technical writing." S18, who works on the Operations side in a Communications role, said she was alarmed to see a recent LLM-generated publication that she felt plagiarised her ideas without credit. She also felt her work was devalued by researchers who did not understand the effort required for science communication writing, saying LLMs led *"to a lack of respect for creative arts or content-generating colleagues."* She also warned that *"it's only a matter of time before research-based content experiences the same [impact from generative AI-authored work.]"*

When we asked managers how they thought generative AI might change hiring, for the most part they said it would change the kinds of skills they looked for in workers but not the total number of workers. Technical managers such as P21 made comments such as, *"I expect that more and more people that we hire will have literacy in AI and machine learning."* P21 continued saying, *"we typically have scientists that run our instruments and collect, acquire and reconstruct data and help interpret it. I suspect we may need a few less of those, but not dramatically. So instead we will need more sort of AI-based people that are able to tweak, let's say large language models."* Non-technical managers were also adjusting the skill sets they looked for. P18, an Operations manager, said *"I would've probably told you 10 years ago written communication was a better skill. Now I'm looking for people who can actually have a conversation on the phone... So I do think that skill sets are going to continue to change."* Overall, there was more concern for non-technical than technical roles, but managers felt that the skills they valued in all roles would be modified based on generative AI's capabilities.

3.5 Discussion

Overall, our findings suggest that participants were familiar with generative AI technology and experimenting with it for a range of tasks from writing reports to automating data science pipelines. Argo usage showed that the number of early adopters for generative AI at the national lab has been growing for both Science and Operations users. These trends, when paired with our interview data, suggest that generative AI tools can add value to science organizations. However, these tools will be the most useful if they are designed to solve real, domain-specific problems, and hurdles such as reliability and accuracy must be improved. In addition, there are risks particular to science organizations that must be addressed. In this section, we discuss design and organizational policy recommendations and ideas for future research directions.

3.5.1 Organization Copilots and Workflow Agents: Design Recommendations and Future Directions

Given the significant overlap between what Science and Operations employees reported looking for in a copilot, we argue a single organizational copilot is needed. However, different domains have unique needs and we see a future where specialized copilots can be designed for different types of workplaces. Based on participant responses, such a copilot could be integrated into organization email services to help workers with tone, with particular support for people for whom English not their first language. In addition, a copilot could have knowledge of common report and grant structures and include pre-programmed prompts to help workers put these together more quickly. It is also critical for a copilot to have access to internal organization data and accurately pull information from academic literature and technical documentation with citations, allowing for interactions such as scientific brainstorming and automatically surfacing correlated data. Moreover, synergy between ideas can advance science more quickly and a copilot could help identify common interests between Science and Operations teams. Lastly, there could be a way for teams to upload team-specific frequently asked questions (FAQs) to the copilot to reduce email burden from common questions. While some of these affordances might be useful in many organizations, science organizations have particular technical context that a copilot must understand for it to be useful.

While organizations could use a single copilot, our findings showed workflows tended to be highly customized depending on the team and project. Organizations that want to support their employees in creating customized workflow agents could provide scaffolding and templates for helping workers with best practices for creating an agent. Organizations could also provide employees with strategies for giving generative AI agents scientific or technical context. Organizations might want to consider a way for workers to share agents they have made with others in the organization. One type of agent we see as being important in science organizations are agents for interacting with and operating complex scientific instruments, an area that merits more research on technical and design considerations for this task. Another type of agent that could be tweaked for both Science and Operations employees are agents for the continuous monitoring and summarization of designated information sources including academic literature and online sources.

Future work could focus on honing in on industry-specific copilot designs, building on prior literature on specific tasks such as writing and programming [114, 160]. For example, answers are needed for what interaction timing is best for organization copilots, how employees can find out about their capabilities, and how complex data sources like science literature can be organized and queried. In addition, future work could look at how to craft a generalised basic workflow agent or template agent that can be adapted by employees in multiple roles and with a range of technical skills. More work is also needed on creating operations workflow agents for common processes around email organization and project management. This builds on literature on AI agents that extends prior to generative AI's development, along with more recent research on generative AI workflow agents in science and elsewhere such as [150, 66, 189]. Lastly, future work can also measure the extent to which generative AI is being adopted in organizations as the technology becomes more developed.

3.5.2 An Organizational Approach to Generative AI Risks

Based on our findings around organizational risks and concerns with deploying generative AI to employees, we outline directions to mitigate these issues in a science organization.

Policies for Publications and Citations

Our findings show that employees have a range of opinions on publication policies and citations when generative AI is used and many were confused about how to proceed. Science organizations, as well as organizations that publish writing to the public, should create clear policies for employees (accounting for all types of roles) around appropriate use of generative AI in writing and citation practices and communicate these policies effectively. While there is no "right answer," organizations can help guide staff so that everyone is working with a shared understanding of what is appropriate use as well as how to cite AI-generated content. Given that we also found significant concerns around reliability and overreliance on generative AI tools (such as hallucinations, lack of scientific or technical knowledge, and overconfidence in incorrect answers), organizations should also prepare for cases where AI-written content includes a falsehood. This may be particularly problematic for science organizations that rely on public trust [123] and grant funding. HCI researchers can help guide what these policies should be and also work on designing technical solutions such as studying how watermarks [155] for generative AI content could work in practice in an organization.

Mitigating Privacy and Security Risks

A key organizational threat, especially for organizations like a national lab that deal with confidential and even classified information, is managing privacy and security with generative AI, a topic that has little research thus far. At this time, many employees are experimenting with both commercial LLMs such as ChatGPT and lab-specific Argo. Participants told us that they are careful and many described redacting sensitive information before placing it in a LLM, however, this means different employees have different ideas about what is acceptable data to share. Science organizations, in particular, need policies around sharing unpublished academic research, research code, and organization-specific information. We also found many participants felt more comfortable, from a privacy and security perspective, using Argo since no queries were stored or shared. This suggests that a viable option is for organizations to make these kinds of internal generative AI systems available for employees in order to mitigate the need to go to outside products. Organizations must, however, make sure their internal options are competitive with external ones since otherwise employees may continue to use external generative AI with better features. Even with an internal LLM that has access to organization data, given employees have different levels of information access, future work could investigate how to design this into the system, particularly for classified information.

Transparency Around Future Hiring and Skills Needed

It is important for organizations to understand that some employees may be concerned about the future of their jobs given the introduction of generative AI. Echoing earlier findings [123], at this time, scientists largely felt their jobs were safe. However, the "AI scientist" project at the lab as well as managers' comments that skills they were hiring for would shift due to generative AI indicate that scientific roles may also be impacted. From an organizational standpoint, we recommend leaders make it clear to employees how they see AI impacting future hiring and what skills will be valued in both operations and knowledge-specific roles.

3.5.3 Limitations

This work provides a case study of a single science organization. Given how little is currently known about organizational use of generative AI assistants, findings from this case study are applicable to a broad range of knowledge work institutions such as those with knowledge specialists (e.g. scientists, lawyers, etc.) and operations workers; in addition, future research should study a variety of science and other knowledge work organizations. When studying usage in the organization, we did not have access to the number of employees accessing commercial LLMs and so our usage data for Argo under-counts total LLM usage. Our participants skewed toward white men, and while this is reflective of the organization a greater diversity of participants might have offered new insights, and future work could correlate specific use cases and risks with demographic features. Our participants also tended to be early adopters of generative AI who were interested in the subject and may not reflect the opinions of employees with little background or interest in the technology.

3.6 Conclusion

In this chapter, I discuss our study of the practical, real-world applications and perceived risks of generative AI use across Science and Operations teams in a multidisciplinary science and engineering research center, Argonne National Lab. To understand current and envisioned generative AI use cases and privacy, security, ethics, and other concerns surrounding generative AI in a science organization, we report on usage statistics for the first release of a private instance of GPT-3.5 called Argo at the lab, a survey ($N = 66$), and in-depth interviews ($N = 22$). We find that there is an upward trend of Argo users, split between both Science and Operations employees, although use is largely experimental at this time. Uses cases fall into either a *copilot* or *workflow agent* generative AI modality. Risks include reliability, overreliance, privacy and security, the impact on academic publishing, and concerns around generative AI taking jobs. We end with recommendations for organizations interested in implementing generative AI systems and for HCI researchers working on crafting these systems.

CHAPTER 4

BUILDING A COMMUNITY-BASED PARTICIPATORY AI SYSTEM FOR CLIMATE SCIENCE

4.1 Introduction

This chapter[1] reports on developing a participatory AI system for climate science. Science-focused AI systems increasingly involve the public, and the success of science relies on public trust in these systems. This is particularly salient for weather and climate AI systems because the forecasts and simulations they produce impact people's decisions about when to evacuate in an emergency, where to move, what kind of infrastructure to install, and more.

One way to build trust in climate AI systems and ensure they benefit everyday people is to include those impacted in the design process. This method of involving stakeholders in the design process for AI systems has been referred to as "participatory AI." While participatory AI appears theoretically promising, the definition is contested and little work exists on how to *practically* go about developing AI systems in a way that is mutually beneficial to AI system builders and the public [58]. There is minimal prior work building real-world participatory AI systems [191]. In HCI there has also been work on "climate services" [157] and tools for citizen science (e.g. [78, 11]). However, participatory design goes beyond the aims of citizen science because it is not just a way to collect data, but to ensure the system has tangible benefits for those impacted. This project seeks to provide lessons on building a participatory AI system in the climate space that provides concrete benefits for multiple stakeholders.

To understand how to build participatory AI for climate in practice, I led a team that embarked on a project that has taken nearly three years. We use this project as a case study to draw practical lessons about the efficacy and challenges of building participatory AI.

1. Collaborators on this project include Thomas Chang, Marshini Chetty, Scott Collis, Nedra Sims Fears, Bonnie Ko, Kanchan Naik, John Rugemalila, and Madison Vanderbilt. Special thanks to all members of the CROCUS team at Argonne National Lab and the Greater Chatham Initiative for their support.

Our research questions are as follows:

RQ1: How can members of the public and scientists each benefit from a participatory approach to climate AI and what are the challenges?

RQ2: What lessons can be learned about participatory AI from building such a system?

We approached these questions in a real-world setting where we collaborated with climate scientists at Argonne National Lab who were developing climate AI models and a community organization called the Greater Chatham Initiative focused on the Chatham community, a region on the South Side of Chicago. Both the scientists and the community organization are part of an existing Argonne initiative called CROCUS. The scientists are planning to study climate and climate impacts in Chicago at a neighborhood scale (as opposed to city scale). In this work, we introduce the term "precision climate" to describe these hyper-local climate models. The scientific approach relies on AI and a sophisticated data pipeline, for example using AI to estimate the depth of flooding on a street from a camera image or to simulate how adding trees to a neighborhood would impact air temperature. As part of CROCUS, Argonne has established partnerships with several neighborhood organizations focused on environmental issues, including the Greater Chatham Initiative. The first goal of the project within this context is to understand needs and perspectives on building climate-focused participatory AI including: 1) understand how Chatham residents view the data collection process and how they might benefit from the hyper-local data collected by scientists, 2) understand how scientists think that community residents can contribute to the climate AI system they are building, and 3) understand challenges to building a participatory climate AI system in practice. The second goal of the project is to co-design a prototype of a participatory AI system with residents and draw lessons from this process.

To do this, we break the methods into four phases. The phases were designed to gain as much information as possible while being unobtrusive to the scientist and community stakeholders. We also wanted to build trust we all sides before taking any action. Therefore, Phase 1 involved relationship building and participant observation where we attended 30+ meetings and 3 in-person science/community events and took field notes. In Phase 2, we conducted interviews with 12 scientists and three community engagement specialists. In Phase 3, we led a series of three in-person "Community Cafes," which were design workshops with Chatham residents (N=62) that followed a human-centered design process to design and prototype an app. In Phase 4, we developed and launched the *Water On My Block* app that focused on flood reporting in Chatham and both benefited residents and contributed to the science AI data pipeline.

By deeply understanding the needs of both climate scientists and Chatham community residents, we have been able to see how their goals could be aligned through this flood reporting tool. On the resident side, Chatham community members can use the tool to report and find out about local street and basement flooding incidents and also use the tool to petition their alderman or the city for resources. On the scientist side, a dataset of flood reports provides insights into climate impact that will improve models and climate simulations.

Project contributions include:

- An understanding of both scientist and community perceptions of a public-facing climate AI system and how the AI system might meaningfully benefit both groups.

- A prototype of a co-designed application that makes climate data actionable for residents and contributes to the AI pipeline.

At a high level, this work will contribute practical insights on how AI systems might be developed in a way that builds community trust through participatory processes.

4.2 Background

4.2.1 From Weather Media To Climate Services

The U.S. initially established the Weather Bureau in 1870 as part of the Department of War [185, 177]. Modern weather forecasting began in the early 1900's, and while the first computer forecast was completed in 1950, computer forecasting only began to be remotely accurate in the mid-1960's [177]. The Weather Bureau evolved, eventually morphing into the National Weather Service in the 1970's and is currently housed in the Department of Commerce [177, 185]. The National Weather Service makes all of its data available to the public and issues forecasts. However, for-profit weather services such as AccuWeather and the Weather Channel compete with the National Weather service while using its data, and can display forecasts that are less accurate because of misaligned business incentives [177]. The U.S. also runs the Hurricane Center, which is focused on forecasting hurricanes [177]. At the time of this writing, the administration is dismantling numerous federal agencies and thus these services may be at risk [25].

Media and communication researcher Marita Sturken coined term "weather media" to describe the spread of weather-related content, first on television and subsequently on the internet [185]. She explains that the Weather Channel was created in 1982 as the first 24-hour channel devoted to weather-related news. As Sturken points out, the Weather Channel's focus on near-term forecasting and prediction meant that longer-term climate trends such as drought were rarely discussed [185, 186]. Sturken describes the increasingly technologically-situated weather reports—such as deploying satellite imagery—as building on a human desire to control the weather via technology. The fact that the weather is ultimately uncontrollable makes it a site of fascination for many.

The proliferation of of mobile phones, social media, and urban sensor networks have led to new kinds of weather media and personalized weather reporting. This has made it a rich area of study for HCI. Sustainable HCI (SHCI) has been a sub-field of HCI since approximately 2007 [28]. A review paper by Bremer, Knowles, and Friday looks at papers over a 15 year period to establish trends in SHCI discourse [28], finding that early work focused on individual behavior change that was later critiqued as being less effective than structural change such as policy intervention. The review paper argues that HCI researchers should use their expertise as designers and engineers to work on "green policy informatics." Example of green policy informatics include a project that rigged stormwater sewers in Detroit with sensors and developed an AI system to manage and automate the flow of water [167], a tool for complex digital services (e.g. Netflix) to measure their carbon footprint [165], and a project to instrument Chicago with air quality sensors and provide QR codes at bus stops so people could view the data [55]. Building on the idea of green policy informatics, other HCI researchers have called for more work on "climate services." Rigby and Preist [157] define climate services as *"the systems that provide climate, weather, and other related information to various stakeholders around to world to inform decision making processes, allowing for planning and preparation for upcoming conditions."* In their review paper, they show a lack of work in HCI on climate services, and the numerous possibilities that exist for important contributions [157]. I situate my work in this chapter within this history of weather media and climate services, approaching these topics from an HCI and design perspective.

4.2.2 Participatory Artificial Intelligence

In an effort to make AI systems beneficial and inclusive, scholars have turned to the long history of participatory design methods. The origins of participatory design trace back to Scandinavia in the 1970s and 1980s, emerging from labor movements that sought to involve workers in the design of workplace technologies [70]. Broadly, participatory design is a design methodology that actively involves stakeholders—particularly end users—in the design process to ensure that the final product meets their needs and values. Recently, scholars have noted a "participatory turn" in the design of AI systems [58] as technology designers grapple with how to make these systems beneficial.

Like participatory design, participatory AI seeks to include the people impacted by AI in its development. While numerous researchers have suggested participatory AI as a mechanism for making AI/ML more equitable and trustworthy, there is not consensus on how to do this in practice [21, 58]. Delgado et al. provide a review of participatory AI practices found across 56 research papers and suggest a framework for precisely describing why participation is needed, who is participating, and what form this participation takes [58]. They define four dimensions of participation: consult, include, collaborate, and own. By sorting papers into these categories of participation, they find that very few papers provide true ownership to the stakeholders involved in the participatory AI design process. In addition, some researchers have warned that involving impacted communities in ways that are performative or even exploitative could cause more harm than good [21, 57]. Participatory AI can be particularly fraught when stakeholders are from diverse backgrounds and have differing needs [57].

While there are concerns about participatory AI, there are also a handful of examples of projects that have benefited communities. For instance, the organization *Queer in AI* has orchestrated a number of initiatives to improve the participation of queer people and perspectives in the design and development of AI systems [152]. In another example, Suresh et al. [191] point out that *"there are not yet examples of how to apply feminist and participatory methodologies from the start... [to] design machine learning-based tools"* and they describe their project that involved co-designing datasets and ML models in partnership with activists working on feminicide. While not explicitly focused on AI, we also draw on work by Sheena Erete who contends that community-based technology design should augment social cohesion, engage small groups within a community, and reflect the interests of the community to encourage participation [71]. In this chapter, I both build on this history of participatory design and contribute practical lessons to the nascent literature on participatory AI.

4.2.3 Citizen Science

This project is informed by prior work on crowdsourced scientific data and citizen science. A citizen science paradigm typically involves non-scientists collecting data and either donating this to existing scientific efforts or using it for their own purposes. Previous work in HCI on citizen science has looked at informatics tools needed to support citizen science [11, 151, 149], technologies for encouraging participation/motivation in citizen science efforts [79, 159], ways of visualizing science information for non-experts [180], and privacy considerations when collecting citizen science data [27].

One example is a paper that looks at a citizen science online community called *Weather-it* [11]. The paper argues that *Weather-it* is a form of "citizen inquiry" rather than citizen science because the online community allows users to develop and carry out their own investigations of weather and share these with others (as opposed to collecting and donating the data to science). In fact, their research finds that while in citizen science projects participants often list contributing to science as a core motivation, *Weather-it* participants ranked this fairly low and instead were mainly motivated by an interest in the topic of weather.

Jennifer Gabrys spent years studying air pollution-focused citizen science projects for her book *Citizens of Worlds: Open-Air Toolkits for Environmental Struggle* [78]. A core barrier she identified for citizen science projects is that when activist groups take this data as evidence to e.g. a government body to advocate for change, the government body can dismiss the data as not sufficiently scientific and thus not believable. Gabrys also explains that even well-collected data does not necessarily translate into action for citizens—it takes significant effort to make this a reality.

While this project builds on ideas from citizen science, I do not label it citizen science because its goal is to be bi-directional. By that I mean that while citizen science implies a single directional flow of information (from participants to science), this project is also concerned with the flow of information from science to residents and ways that this information can directly benefit residents.

4.3 Methods

4.3.1 Research Context

The CROCUS[2] project at Argonne National Lab is an "urban integrated field laboratory" where climate scientists are deploying approximately 21 multi-sensor instruments around Chicago to measure climate/weather and climate impacts at a hyper-local scale. Climate has traditionally been measured at a city scale, often with a single weather station for an entire city or town. CROCUS hopes to shift this paradigm by collecting neighborhood-specific climate measurements (e.g. temperature, air quality, etc.) and climate *impact* measurements (e.g. using machine learning to extract the depth of water on a street from image data).

The output from the climate scientists' work on CROCUS will be raw data, accessible using a Python API, and computational models of neighborhoods. CROCUS has aimed to partner with local community organizations. Upon learning about CROCUS, I led the development of a collaboration between human-computer interaction researchers and the CROCUS team with the goal of bridging the gap between the scientists and community residents such that residents could benefit from, and participate in, the climate data/AI system. We ultimately developed a deep partnership with one community organization—the Greater Chatham Initiative (GCI)—focused on driving economic growth in the Chatham region of Chicago.

Chatham is a neighborhood on the South Side of Chicago. 94% of Chatham residents identify as Black and Chatham's median household income is $39,348 [2]. Chatham has a long cultural history, for example it was the home of famed gospel singer Mahalia Jackson. The Greater Chatham Initiative "was born from a comprehensive plan for economic growth

2. https://www.anl.gov/crocus

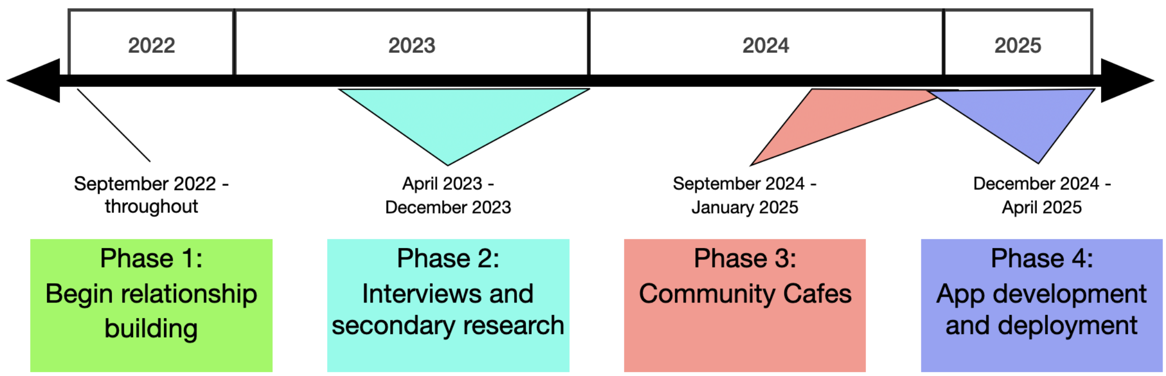


Figure 4.1: Timeline for project, include the four phases: relationship building, interviews, community cafes, and development and deployment.

and neighborhood vitality."[3]

4.3.2 Phase 1: Relationship Building and Participant Observation

Prior to beginning formal interviews or design activities, we spent considerable time building relationships with science and community stakeholders. Our aim was to understand as much as we could about the context of each stakeholder with as little burden as possible given their busy schedules.

We conducted participant observation by attending 30+ meetings and events and taking field notes (beginning September 2022). These included virtual science research meetings and community organization meetings. They also included an in-person event celebrating the deployment of one of the measurement instruments, a community-organized event focused on flooding where we did a 30 min presentation with discussion, and an in-person conference bringing together scientists and community members. In addition, we reviewed community organization websites and public statements and followed the main Slack channel used by scientists. Through these activities we developed trust with each stakeholder, learned broadly about their interests and priorities, and showed that we were committed to a longer-term

3. https://www.greaterchathaminitiative.org/about-us/

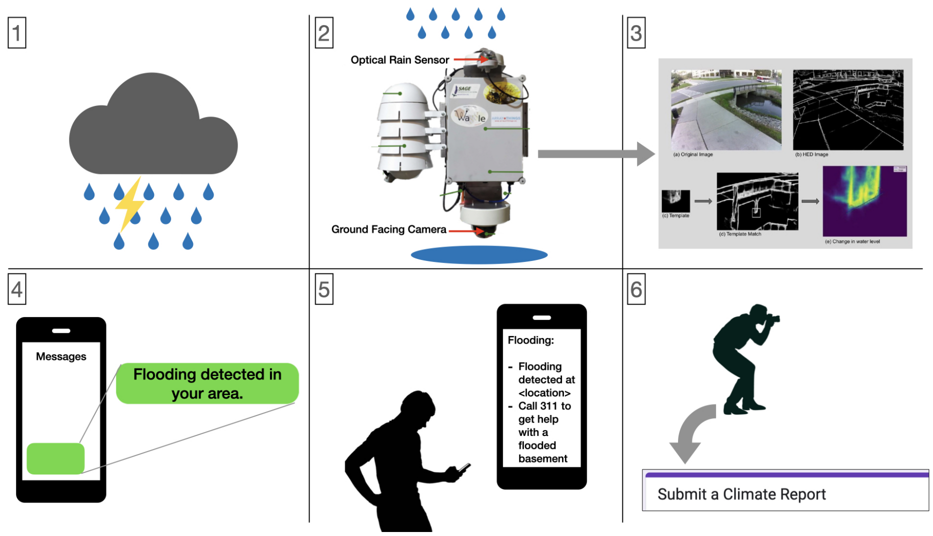


Figure 4.2: One of several original storyboards shown to community partner representing a flood alert text system.

engagement.

4.3.3 Phase 2: Semi-Structured Interviews

I conducted 15 one-on-one semi-structured interviews with scientists, community members, and community engagement strategists between April 2023 and December 2023. Interviews lasted 30-45 minutes and were done virtually over Zoom. I recruited participants by reaching out directly to members of the CROCUS initiative via email and Slack, as well as CROCUS partners.

Our objective in the interviews was to understand stakeholder goals for the CROCUS initiative and broader climate goals (e.g. *How would you explain the main goal of CROCUS more broadly? What do you think are the current or future impacts of climate change on you personally?*). When talking to scientists, we wanted to understand their perspective on community involvement in developing and consuming climate models (e.g. *Do you have any ideas for how community members could be involved in the CROCUS AI pipeline? If you could pick any CROCUS data to be put in a visualization for non-experts, what would it be*

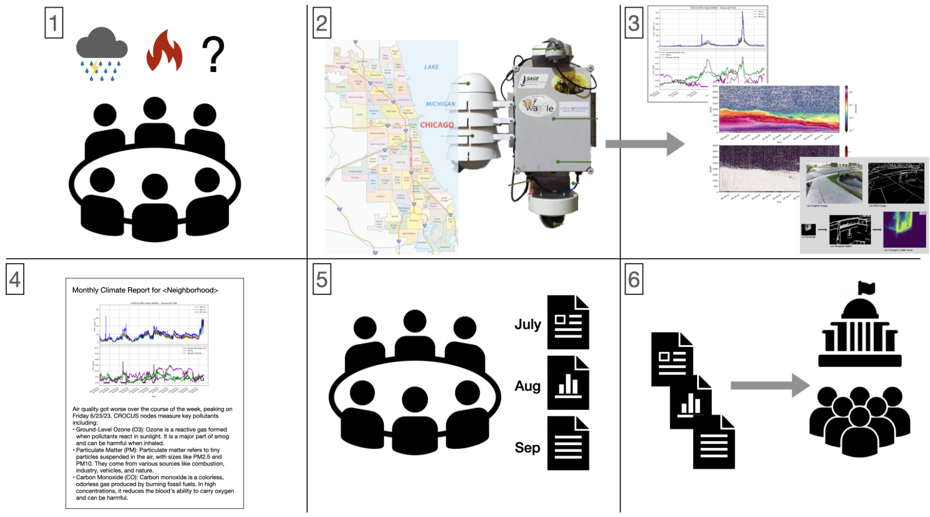


Figure 4.3: One of several original storyboards shown to community partner representing community reporting.

and why?). When talking to community engagement specialists, we also wanted to understand how they thought community residents might get involved (e.g. *What are examples of community partnerships that have worked well in the past and why? From your perspective, what do the community partners want to get out of working with CROCUS?*).

Interview participants were predominantly on the science side and included both researchers and software engineers. Participants working on the science side were mostly white men with doctoral degrees between the ages of 25 and 44, which is representative of the lab population. Community engagement specialists identified as Black or African American women, and all had a long history of working with communities in the Chicago area on a variety of projects. See Table 4.1 for the full list of participants.

Table 4.1: Interview Participant Demographics

ID	Gender	Age	Race/Ethnicity	Education	Occupation	Years in research*
P1	M	45-54	White	Doctoral degree	Atmospheric Scientist	25
P2	M	25-34	White	Master's degree	Research Software Developer	2
P3	M	25-34	White	Master's degree	Atmospheric Scientist	3.5
P4	M	25-34	White	Master's degree	Research Software Developer	7
P5	F	35-44	White	Doctoral degree	Atmospheric Scientist	8
P6	M	35-44	Middle Eastern or North African	Master's degree	Research Software Developer	7
P7	F	35-44	Asian	Doctoral degree	Computer Scientist	5
P8	M	35-44	Hispanic, Latino, or Spanish origin	Doctoral degree	Computer Scientist	3
P9	M	35-44	Asian	Doctoral degree	Atmospheric Scientist	10
P10	M	35-44	White	Doctoral degree	Atmospheric Scientist	7
P11	M	35-44	Asian	Doctoral degree	Computer Scientist	4
P12	M	55-64	White	Doctoral degree	Atmospheric Scientist	32
P13	F	45-54	Black or African American	Master's degree	Community Engagement Specialist	N/A
P14	F	65-74	Black or African American	Master's degree	Community Engagement Specialist	N/A
P15	F	55-64	Black or African American	Bachelor's degree	Community Engagement Specialist	N/A

*Number of years in research after terminal degree.

4.3.4 Phase 3: Community Cafes

In collaboration with our non-profit community partner, we conducted three "Community Cafe" events with Chatham residents. Community Cafes are a method used by our partner to engage residents during an in-person discussion-based event, and they build on the World Cafe model [1]. Aspects of the World Cafe method include: a welcoming environment, small group discussions at tables, and providing food to participants. We provided lunch at all our events and compensated participants with an Amazon gift card for their time. All events were held in-person in Chatham. Participants were recruited by our community partner, the Greater Chatham Initiative, via their email listserv.

Our three Community Cafe events followed the human-centered design process. Our first event was in September 2024, it lasted three hours, and had 21 participants. In the first event we focused on understanding Chatham's climate concerns broadly, and how residents thought technology might support or address these concerns. We led discussions with questions such as *What are some examples of times when a climate issue was a problem for you?* and had participants do a drawing exercise where they created a storyboard for a new app in a climate scenario of their choosing. See an example of storyboards in Figure 4.4. In addition to discussions, we had participants fill out a survey providing feedback on their main climate concerns and demographic information.

Our second Community Cafe event was in October 2024, lasted 45 minutes, and had 27 participants. It focused on refining what the app should do and how people should interact with it, building on findings from Community Cafe 1. We provided three possible mock-up app designs to participants and had them vote and provide feedback on these. We emphasized that the final design could combine different aspects of each mock-up. The first mock-up was for a system that would send text messages about air quality and allow users to text back to develop personalized AQI (Air Quality Index) thresholds for poor air quality; the second mock-up was a web app for reporting local flooding incidents; the third mock-up

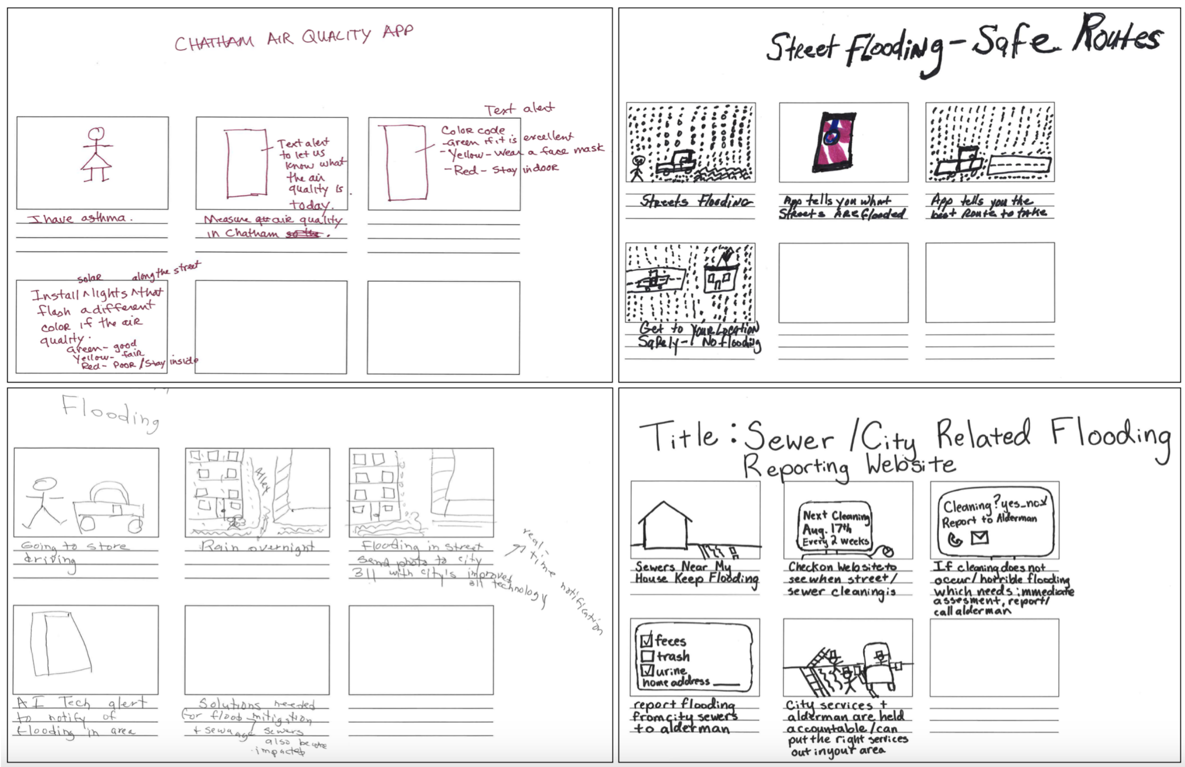
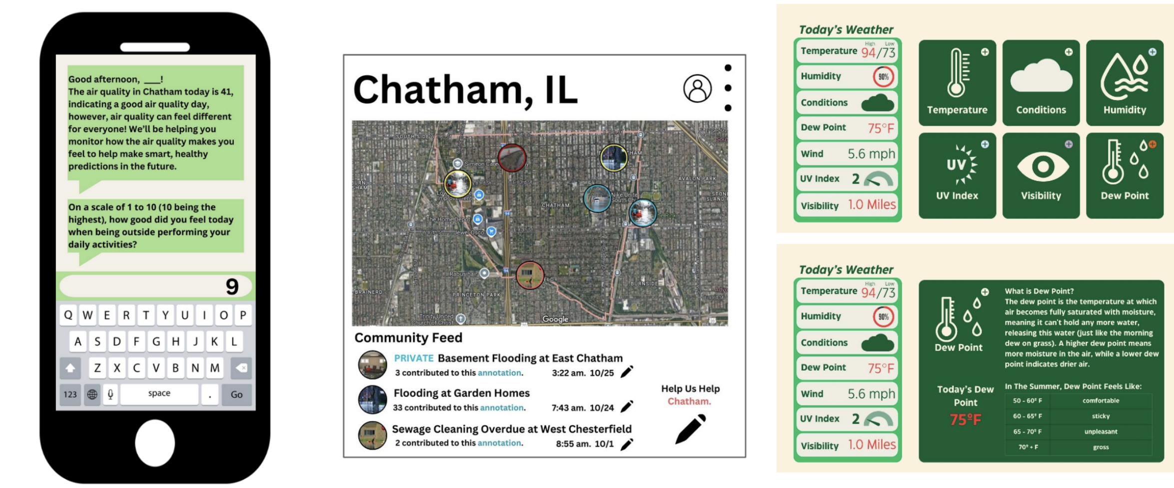


Figure 4.4: Storyboards created by participants in Community Cafe 1 showing how an app could support different climate scenarios.



Mockup 1: Air quality texts **Mockup 2: Flood reports** **Mockup 3: Educational website**

Figure 4.5: Mock-ups designed by our team presented to community residents during Community Cafe 2 for voting and feedback.

was a website with Chatham-specific weather/climate resources. See mock-ups in Figure 4.5. In this event, we had participants draw directly on the mock-ups to provide feedback, and also had them fill out a survey which included voting on their favorite mock-up.

Our third Community Cafe was in February 2025, lasted three hours, and had 14 participants. Based on feedback from the prior events, we built a prototype of a flood reporting web app and in this Cafe we focused on getting feedback on the prototype. We brought laptops and had participants try the prototype live. We conducted a discussion around usability and features as well as on privacy and data governance given the sensitive nature of user-generated flood reports. We had participants fill out a paper survey during the event.

4.3.5 Phase 4: App Development and Deployment

We began developing the flood reporting app in December 2024. We waited to begin development until after we had analyzed the data from the first two Community Cafes in order to incorporate all community feedback. We chose to build a web app, as opposed to an iOS or Android app, because it would be more flexible and we did not have the engineering resources to develop multiple versions of the app.

The web app is designed to provide an interactive experience for tracking flooding incidents in Chatham. The platform uses a combination of modern technologies to ensure smooth performance and secure data management. The backend is powered by Node.js and Express.js, with MongoDB managing all data storage, including user-uploaded images. The frontend is built with React.js, and to enhance mapping features, the app uses Leaflet.js alongside the Google Maps API for location searches. Axios facilitates communication between the frontend and backend. User authentication is handled through Google Firebase for secure access, and the entire application is deployed via Amazon EC2 for reliable hosting.

4.3.6 Analysis

Interviews were transcribed by Rev.com under a non-disclosure agreement. We qualitatively analyzed the interview transcripts using the software MAXQDA in two rounds using thematic analysis [162]. First, we did an initial coding round using a codebook developed based on the interview protocol (see codebook in Table 4.2). Second, one member of the research team developed subcodes for each top-level code. Then, a second member of the research team reviewed all subcodes. For each top-level code, one member of the research team developed a thematic summary of the code based on the subcodes. In this thesis, we focus on the top-level codes bolded in Table 4.2.

We also conducted thematic analysis on the participant observation field notes, surveys from the Community Cafes, and sketches and marked up mockups from the Community Cafes. After each Community Cafe, two members of the research team qualitatively coded hand-written sketches and feedback, survey responses, and field notes using MAXQDA. We then summarized these codes into themes representing participant feedback and used this thematic feedback to make design decisions for the next iteration of the app.

Table 4.2: Codebook for Interviews

Code	Code Description
Individual goals	Individual goals each interviewee has for their own work
CROCUS goals	How interviewee understands the broader goals of CROCUS
Climate concerns	Climate concerns that the interviewee has broadly
How CROCUS can contribute	What do interviewees see as the main contribution of CROCUS to community partners
How community can contribute	What do interviewees think that communities can do to help with
Stakeholders	What other stakeholders may be interested in CROCUS data
Perceived gaps	What do interviewees perceive as gaps between scientists/communities
Climate actions outside work	What climate actions have interviewees taken outside work

4.4 Findings 1 (Phases 1–3): Understanding Community and Scientist Needs and Challenges for Participatory Climate AI

In this section, we report findings from the first goal of the project to understand the needs and perspectives that both community members and scientists had with respect to a participatory AI system for climate science. In Section 4.4.1 we detail how community members might be able to use and benefit from climate AI and precision climate. In section 4.4.2 we explain how scientists said they could benefit from community participation in their AI system, particularly data collection. In section 4.4.3 we examine challenges to bridging the gap between communities and scientists when setting out to build a participatory climate AI system.

4.4.1 Community Uses for Climate AI

First we sought to understand how community members might want to use precision climate data. We found that while numerous community members hold deep expertise in local climate issues, quantitative precision climate data could be used to reinforce this knowledge and advocate for resources. When speaking with community leaders and members about how precision climate could be used in individual and community decision-making, we found interest in quantifying "bang for the buck" in green infrastructure investments, real-time alerts, and severe weather prediction.

Augmenting Existing Community Knowledge and Advocacy

Chatham community members have been advocating for environmental issues in their neighborhood for many years. Community leader P14 spoke to a group of scientists and explained that due to excessive flooding in her neighborhood, while she was growing up her basement flooded numerous times and caused two electrical fires. She has no memorabilia left from childhood because of these events. This neighborhood sits at the intersection of two water reclamation systems in the Chicago area. When it rains hard, these systems fill from the center to the periphery and sewers can overflow directly into the neighborhood, flooding basements and streets. In addition, the neighborhood sits near pollution-generating sites such as a highway that lead to poor air quality. Historically, the neighborhood has not received the funding necessary to tackle these problems. For example, Chicago has a program where residents can call 311 if they need help with infrastructure issues such as a pothole or flooded basement. Community leader P14 explained that most residents will not call 311 due to the fact that historically the effort to report has not been met with concrete help. In the Community Cafe surveys, participants made comments such as, "*311 operator responded with lack of concern*" and "*follow-up is terrible*." Some residents also worry that the data collected by 311 might lower the value of their homes.

While knowledge about these types of climate issues is deeply embedded in the neighborhood, funding bodies such as the city or grant-giving institutions often seek quantitative data to back up the claims. P14 explained the need for precise neighborhood-level data on issues such as air quality saying,

> *"I really like the whole notion of it being hyper-focused so that we could determine how [the neighborhood] does or does not relate to the city of Chicago... I speculate that our air quality is probably worse than what's reported, generally, given [proximity to transit hubs], so it would allow us to know that a good air quality day in the city of Chicago is a fair one in Chatham."*

Advocacy is a community-level application of precision climate data, meaning it is organizations within the community that can take action on the data (as opposed to individuals). This use case is also aligned with the kind of data scientists are used to generating for their own grant applications.

Climate-Related Infrastructure Investment

Community members also wanted to use precision climate data to help decide which climate adaptation strategies and investments would be most worthwhile. For example, there are a number of interventions that might soak up water and reduce flooding such as planting trees, installing bioswales, utilizing rain barrels, and updating sewage pipes. Residents were interested in both community-wide investments (such as fixing the sewage system) as well as individual-level investments (such as adding rain barrels or sump pumps to homes). Community leader P14 summarized saying,

> *"Success would be having an action plan where we can look at what activities have the biggest bang for the buck [...] That we have documented that if we change where the sewer connection pipe is, or we plant 10,000 more trees, that we would have an action plan to go ahead and do that."*

The concept of most "bang for the buck" came up numerous times when describing how to use the data. The community was interested in data that supported which of these interventions would be most effective locally—particularly for limiting flooding. As P14 said in a public statement, *"[precision climate data] can help us make informed decisions because we have limited resources. We don't want to guess, we want to know."*

Real-Time Alerts

Community members emphasized the relevance of targeted and precise real-time alerts, particularly for individual decision-making. One example that came up multiple times was deciding which block to walk on in the case of poor air quality. Residents remarked that their neighborhood bordered a highway and that this could be a source of unhealthy air. More specifically, community leader P14 wondered about what route children should use to walk to school saying, *"If a kid's got to walk in poor air quality, what does that look like?"*

P14 said that an alert function would be important since not everyone checked the weather on their own, and that the alert should be paired with information on how to act:

> *"I'm frankly shocked at the number of people who don't check if it's a good or bad air quality day... We speculate, just based on the number of AC units that we see attached to apartment buildings that have five or more units, that maybe 20% to 30% have air conditioning. So to help people visualize the call of actions needed, so they're willing to put something in place."*

One Community Cafe participant also mentioned that it would be nice to have an alert for particularly good air quality, so she could garden or go on a walk. There was interest in the real-time alert coming in the form of a text message as well as interest in a map visual. P13, a community engagement specialist said in relation to air quality,

> *"[I'd like] an aerial landscape or map, which shows higher pollution areas... or I don't know if you can get a detail with block by block, which would be amazing... Just so people know if they're walking into something."*

Multiple Community Cafe participants also said real-time alerts would be helpful for flooding. Participants said this could be an alert about the start of heavy rainfall that warned residents to prepare for possible basement flooding. P14 noted that real-time alerts would also be helpful for flash flooding incidents, explaining that while driving it was possible to be caught in a flash flood and it would be useful to know which roads to avoid.

Severe Weather Prediction

In addition to real-time alerts that would tell residents when a severe weather event had already started, there was significant interest from community members in being able to predict such events in advance. When our research team gave a presentation at a flooding-focused event in the neighborhood, the first point of discussion was whether flooding could be predicted, and thus planned for and possibly even prevented. The interest in prediction was particularly salient for precision climate since residents already have access to city-wide forecasts. Community members felt, however, that the city forecasts were often not applicable to them and wanted a neighborhood-level, and ideally block-level, prediction. As we describe in Section 4.2, prediction is a particularly challenging scientific and communication problem.

4.4.2 How Scientists Envision Community Could Participate in AI Pipeline

Overall, scientists focused on how community members could contribute data and insight that would shape the climate models simulating their neighborhoods. Scientists differed, however, on what type of data they thought would be most useful, how it could be collected, and how it could be used in the AI system.

Types of Community-Sourced Data

Multiple scientists spoke about images as a useful form of data, ideally with categorical labels. Images could show both flood and snow depth that could be extracted using machine vision algorithms. Categorical weather labels overall were of interest, such as to describe current conditions as "icy" or "snowing." Scientists also said collecting geolocation was critical and that this needed to be at least as precise as a city block; for their purposes, a zip code would be too broad.

How to Collect Community-Sourced Data

Scientists brought up examples of meteorologists using social media to get real-time community data via hashtags, and that this could be a possibility. They also mentioned lower cost sensors that residents could place and monitor. Lastly, they said a dedicated app to collect information could work well. Scientists said it would be important to train residents on how to collect scientifically useful data, such as what a "good" photo would be or how to use the sensors.

How Community-Sourced Data Could be Used in AI Models (Or Not)

One possibility participants mentioned would be training models directly on the data submitted by community members, but opinions on whether this would work were mixed. One scientist said '*'Because the big thing in AI is having high quality labeled data sets. So what I would love to be able to do is actually have people report urban flooding throughout Chicago, and then we could train a model with rainfall as input and inundation reported by people as an output."* (S1). However, another scientist did not think the data would be high-quality enough saying, *"If you want to have a network that can explain something... you have to have a continuous flow of data with a certain quality... Otherwise, the network will cluster [varying quality data] differently."* (S8).

Another way to incorporate community data would be through model validation and testing. This was the most popular use of community data shared by scientists. S2 described how it would be important to include a human side to model evaluation saying, "*Let's say we have a big heat wave and people mentioned that the Woodlawn area was quite a bit warmer... Not just saying that, 'Oh, the RMSC was 2.5.' Qualitatively going through and being like, 'Okay, it looks like Woodlawn was warmer than Bridgeport.' Some of these things that are more important to the human aspect and not just... our typical diagnostics that we use when we're going through and evaluating models.*"

The final way that scientists said they could make use of community input was as a way of making sense of AI model outputs. For example, S12 said there could be unexplained spikes in the data without an apparent cause:

> *"But occasionally with our different sensors on the roof, we'll get this isolated spike in carbon monoxide or something. In addition to being able to alert people*

nearby, [it would be helpful to ask] 'Hey, everybody look around. Are there any big trucks? Do you see any smoke plumes?'... I could see people wanting to engage with that, just like there's storm spotters and pollution spotters."

In addition to unexplained spikes, other scientists mentioned distinguishing between categories that instruments have trouble with, such as the difference between dust and pollen which are a similar size. Lastly, community residents could help with processing complex data, for instance identifying cicada noises that an AI model might struggle with naming.

4.4.3 Challenges to Community-Based Climate AI

In this section, we draw on interviews with climate scientists as well as community members to identify gaps that must be bridged in order for precision climate to move from the lab to the community. First, despite the fact that scientists were excited by the prospect that community members could use the data they carefully collected, there is a lack of both social interfaces (between scientists and the public) and technical interfaces (between the raw data and a more user-friendly format). Second, there was a tension between residents' desire for high-accuracy and hyper-local prediction/real-time alerts of climate events (particularly severe ones) and scientists' tendency towards preference for a carefully curated backwards look at data since given tricky questions of accuracy and conveying uncertainty. Third, the need for trust between scientists and communities—and the significant time it takes to build this trust—is critical for precision climate applications to be successful.

Lack of Sociotechnical Interfaces Between Scientists/Raw Sensor Data and Community Members

The scientists we interviewed were overwhelmingly excited about community members who were not professional researchers using the data they produced. P2, a software developer on the climate science team, said that he hoped:

> *"The people would be looking at the data all the time, especially at the community level... So we post events looking at after a storm has gone through or after there's been a big heat wave during the summer that people look to [our data] for the information there... And really just seeing the data used by the... community partners and the Chicagoland area, seeing it in the hands of community members."*

P4, another atmospheric science software developer, echoed this sentiment saying, *"I would like to see large scale community involvement in actually using the data we're producing."* Despite this excitement for using the data, however, no stakeholders fully grappled with the social and technical reasons community members were not equipped to seize the opportunity to work with the raw data.

We argue there are a lack of "sociotechnical interfaces" because we found that both social and technical interventions would be needed to realize scientists' goals of having the community *"looking at the data all the time."* On the technical side, the data is stored in several databases and requires skills such as using an API, Python, and reading CSV files. Even with this skillset, conducting data science analysis is time-consuming. Community volunteers were already overworked and likely did not have time to dive deep on the raw data. Improving the technical interfaces, however, must be paired with improved social interfaces. Scientists are already seeped in the *context* surrounding the raw data. It makes sense that, like fish in water, it is easy to forget this context was learned and must be shared with community members for them to engage with the data in the same way.

Tension Between Residents' Desire for Real-Time Data and Scientists' Concerns About Data Quality and Uncertainty

In Section 4.1, we described how residents wanted a way to precisely predict severe weather events that would personally impact them. This need is in tension with scientists' desire to produce highly accurate data with error bars that measure uncertainty. For example, on a particularly foggy day scientist P1 noted in their research Slack channel, *"Wonder if the fog is playing havoc with our [particulate matter] measurements today"* and showed a graph with higher than normal measurements. After some discussion, the group of scientists concluded that there was not a spike in pollution but in fact the increase was due to dense fog. Had a real-time air quality alert been deployed to the public, it would have needed to subsequently be retracted, likely damaging public trust in the data.

Prediction is even more uncertain than real-time alerts, given no one knows for sure what the weather will be in the future. Conveying this uncertainty—which is particularly high for precision climate—is a challenge. Several of the scientists we spoke with were particularly hesitant to endorse prediction as even a possibility for precision climate. P3, an atmospheric instrumentation specialist, summarized this perspective saying:

> *"The flooding information is going to be something that's years down the line. [Residents] are not going to see some real time,"Hey, it's about to flood for you." They won't ever get that. It's going to take us forever to develop something like that."*

A couple thought precision prediction *might* be possible, and that real-time alerts were well within the realm of possibility. P9, a researcher, summarized saying:

> *"It'll also give [residents] real time guidance during extreme weather events. So I think that they can consider as a benefit and once you get more and more data from these kind of nodes the algorithm can become smarter and they can give much better guidance. And if not prediction I would say it would be telling you how bad your weather is today and how much care you should be taking."*

Others said they would be more comfortable providing explanations and public-facing write-ups of events that had taken place in the recent past. Backwards-looking reports, however, were of less interest to community members given they did not aid decision-making in the face of an extreme weather event—in other words, they were less actionable.

Differing Understandings of Surveillance, Community Benefit, and Trust

Both scientists and community members were aware of the need to build trust. We found, however, that each group emphasized slightly different viewpoints when describing how to build trust that we surface in this section.

Scientists were primarily worried that the sensors they needed to place in neighborhoods would be interpreted as surveillance devices, particularly in areas that had been over-policed. P3, an atmospheric instrumentation specialist, voiced this concern which we heard multiple times, saying,

> *"[The community is] going to see something that has a camera on it, potentially two cameras and be like, 'What is this surveillance device doing?' So that's where I'm concerned on the community reception is if we have a camera going at all times and we say, 'Hey, it's for science,' they're going to just think we're studying them when it's not studying them. It's looking at the vegetation, the roads, the sky."*

Multiple scientists said they felt it was important to engage with community members to explain that the sensors were not spying on them. For example, atmospheric scientist P1 said scientists needed to,

> *"Help communities accept these tools, these important tools, into their neighborhoods... the answer is... critical data literacy. And then I do think it's really important for [us] to have a robust public outreach component. Lots of listening, lots of understanding the concerns of the communities, and doing things to address them."*

To be clear, community members do care about privacy. Their view of privacy concerns was even more expansive than scientists.' For example, while scientists were focused on invasion of privacy via policing, community members surfaced additional privacy concerns such as fear of housing prices going down if data about the extent of flooding was made public.

But while community members care about privacy and policing, the most significant concern we heard was that the science concretely benefit the community. In other words, having privacy-preserving technology was necessary but not sufficient. In addition to being over-policed, this community had been over-studied but had not seen direct benefit from most research done on them. Managing recurring basement flooding and its mental and physical impacts had also taken a significant mental toll on the community, as described by numerous Community Cafe participants. P15, a community engagement strategist summarized saying, ‘*‘If you're going to pick [at these communities] and you're not going provide any compensation or any solution, then it's really injurious at this point."* In a public statement, community leader P14 described the ideal way of working with scientists while seeing a "tangible" result saying,

> *"Equity looks like when the decisions are driven by the community for the community. We're telling you what's good for us and then we're working collaboratively to address the issue. I find that when we talk from a scientific lens, not based on 'please help us,' but we have the science that can inform, that if you change where that sewer pipe is and move it, about 50% of the basements will flood less. That is something that is tangible."*

P5, an atmospheric scientist who had spent time with community members reflected:

> *"I think having worked with the community partners during this past year, hearing their side of the story, I see that there is this disconnect between the communities and the scientists. The scientists get excited by the science they do... But me personally, I never looked beyond that how it is going to be impacting somebody else. So now it has become kind of personal because we know those people, we have heard their stories and we have seen their discontent and distrust a little*

bit as well because they know we exist, we do our science, but how is it going to impact their well-being, their neighborhoods? That part is the disconnect."

This quote shows how she developed an understanding over time that the key for building trust with the community was much more than addressing privacy concerns, but was actually "impacting their well-being." As she states, this is a challenge scientists are not accustomed to addressing.

4.5 Findings 2 (Phases 3–4): Building the *Water On My Block* App

4.5.1 Key Design Decisions Based on Community Feedback

We used an iterative co-design approach to develop the app, including the decision to make an app in the first place (see Section 4.3 for more details on methods). Ultimately, we developed *Water On My Block*, a web app that works on both desktop and mobile (see the screenshot in Figure 4.6). *Water On My Block* lets users report a flood incident, view flood incidents on a map and in a community feed, and see resources for flooding specific to Chatham. When reporting a flood incident, users are asked questions about the flood such as location (basement or street), water depth, and how long the standing water has been there. Our partner organization has final say on who the data is shared with, but has agreed to share it with the climate science group we have been working with to improve flooding models of the neighborhood. After starting with a blank slate, the iterative co-design methodology we used led to a number of key design decisions that we outline below that culminated in this app.

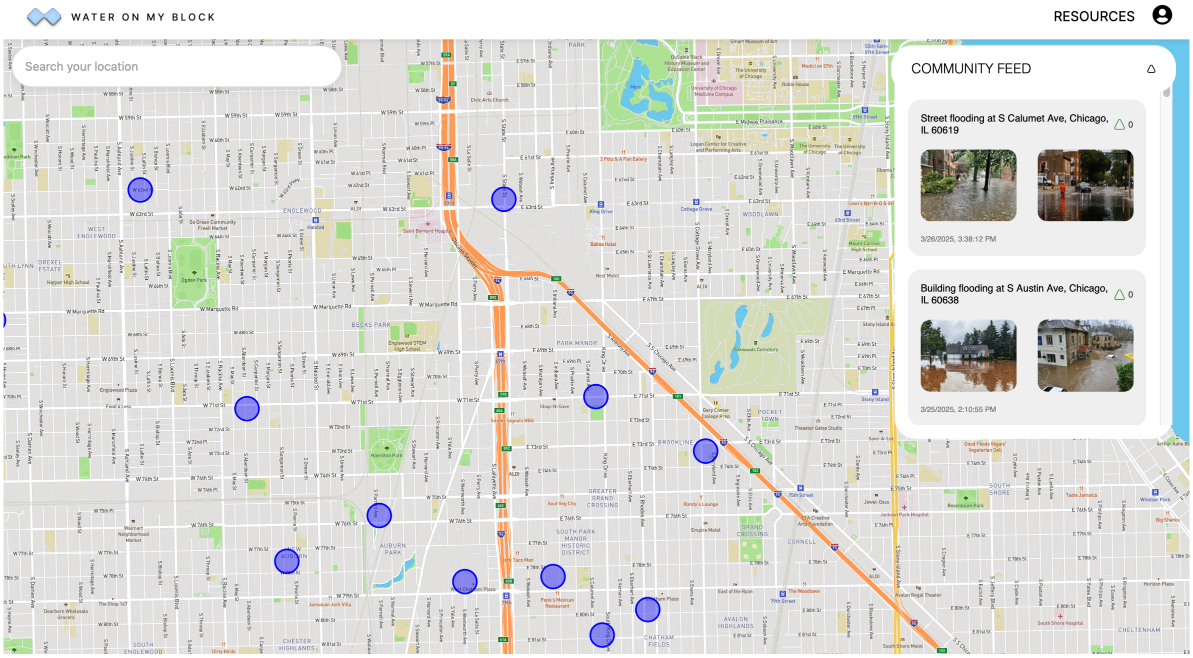


Figure 4.6: Screenshot of the *Water On My Block* web app. Circles show where there is a flood incident. We do not show exact addresses to maintain privacy.

Deciding to Build An App

The first decision was whether to build an app at all. During early talks (prior to the start of the formal research process) with our community partner, we provided some ideas of possible apps via storyboards. However, we also provided more socially-focused ideas including a Data Ambassador program where we would provide data literacy training to community residents and structure dialogs between the Data Ambassadors, residents, and scientists to include the community in the science and AI development process. Our community partner provided feedback that they would prefer us to build a tangible app that could be shared with all community residents as opposed to something like the Data Ambassadors program. They felt that having something tangible and shareable would better convey the importance of the science to more residents. An app was also less burden on the community, since it required less time-intensive involvement than a training program. Another limitation of a training program was our inability to provide a formal educational certificate, which again would have been something tangible to show the community. Separately, structural interventions such as retrofitting buildings or planting trees would have been helpful but was not within our team's skill set or budget. Thus, given our skills and computer scientists and HCI experts and the community's desire for something tangible related to climate we decided to build an app.

Focus on Flooding

In our first Community Cafe, we asked residents via a survey what climate concern was most important to them. Flooding and air quality were both significant concerns. We also asked participants to storyboard ideas for how a climate app might be useful to them, and many of the storyboards revolved around these concerns. In the second Community Cafe, we presented three possible app designs, based on the feedback from Community Cafe 1, and had participants vote and provide feedback. One design related to flooding, another related to air quality, and a third was a more general educational website. The most participants voted on the flood reporting app. In addition, scientists had expressed interest in knowing about flooding incidents that their instruments could not track (e.g. basement flooding). It is important to note that we factored this in secondarily to the community feedback, but given the community's interest in a flooding app we realized it could satisfy the goals of multiple stakeholders. Thus based on the combined data from Community Cafe 1 and 2, we decided to create a flood-focused app.

Interface: Website, Map, and Community Feed

During Community Cafe 2, we presented two different core interface options: a text message interface and a web interface. While one was shown for air quality and the other for flooding, we explained to participants we could make either interface work for either climate concern. Several participants described that they already get too many text messages and it would be overwhelming to get more, with one participant saying the Community Cafe 2 survey that this would be "*information overload*." The map interface with a community feed was appealing to people because it mirrored existing experience with social media websites. In addition, the map could be useful during extreme weather events for the community to share information. Lastly, participants brought up that it would be important to have a tool usable by seniors since "*they pay closer attention to what's going on in the community*."

Custom Resources Page for Chatham

While participants in Community Cafe 2 were primarily interested in a flooding app, there was also considerable interest in access to educational resources related to flooding. We discovered through conversations with participants and the community leader that these included background on topics such as weather and climate concepts (e.g. there was confusion around what "dew point" is), what to do in the event of a flooded basement, contact info for local plumbers and others who might help with flooding, flood insurance info. To address this unmet need, we added a page in the web app called "Resources" formatted like a FAQ to provide this information.

Aiding Advocacy Efforts: Aldermen and 311 Connection

In addition to providing benefit to residents via the app interface, based on resident feedback we also wanted to make the data actionable for advocacy. To do this, we included a link to auto-generate an email with the flood incident data for the local aldermen (local government official). While reporting to 311, a city of Chicago reporting service, was not always seen as useful we also added a link to this form in case residents also wanted to report the flood there.

Data Privacy and Governance

Data privacy and governance have been a top priority since the beginning of the co-design process. We decided early on that any data collected would be owned by our community partner and they would have sole discretion when decided who to share the data with. In Community Cafe 3, we had an extended discussion on data privacy with participants, particularly in terms of the interface design. Residents had mixed opinions on if they felt comfortable sharing flood incidents on a public map, particularly for basement floods. To address these concerns, we made a number of interface decisions. First, we do not have addresses listed on the public interface and show flood incidents using a circle the size of a neighborhood block. We add random noise to where the circle is centered so it cannot be inferred from the circle. We store exact addresses on the backend so that these can be used by the community partner and scientists. Second, we allow anyone reporting an incident to do so privately so it does not show up on the map at all. We also allow for the incident to show up, but to hide the photos in case they are identifying.

4.6 Discussion

The initial goal of this project was to design and develop a participatory AI system. As noted by Delgado et al. [58], while many papers claim to follow a participatory process when building AI, most merely allow stakeholders to consult rather than provide stakeholders with true ownership. A test of true ownership includes the ability to stop the AI project from being built at all. This means that, in fact, starting out with the intention to develop a participatory AI system prior to consulting stakeholders turns the term into an oxymoron. We realized that once we engaged with our stakeholders, particularly the community partner, and wanted to provide them with full ownership over the project, then we needed to let go of any particular technical outcome we originally wanted. Once we began discussions with the community partner and residents, we realized their focus was on developing what has been called a “climate service” [157]. To the community, whether this climate service involved AI or not was beside the point—what mattered is that it directly addressed community climate concerns. For the scientists building the AI system, however, it was exciting to be able to include community data in their models (or in the model testing). So did we build participatory AI? From our perspective and the perspective of the climate scientists, yes. We created a system that adds community-sourced data to the AI pipeline and engaged community members in the scientific process. But from the perspective of the community, we created a climate service.

Given that the app we produced (*Water On My Block*) functioned both as a participatory AI system for climate scientists and as a climate service for the community, addressing the needs of each group but without requiring consensus on the ultimate goal of the app, we argue that the app functions as a boundary object. Boundary objects are artifacts that have interpretive flexibility, meaning the same object can be used/understood by different groups in distinct ways, and they also “*allow different groups to work together without consensus*” [183]. Throughout our discussions, interviews, and Community Cafes that nominally centered around the design of this app, we also provided ample opportunities for scientists and community members to learn about each other and find ways of collaborating and communicating effectively, while also each working towards their own goals. Thus we contend that the participatory AI design process can be about developing a sociol interface as much as a technical one: the goal of crafting a technical artifact can allow for the mutual shaping of stakeholders during the process that has implications outside of the final object.

We have shown in this project that over the course of multiple years, with a highly targeted project, where the AI is a relatively small-scale machine learning system, we could align stakeholders in a participatory design process. But this begs the question as to whether this kind of method for developing AI can scale. To this end, we propose the concept of “community data partners.” AI system builders can partner with communities to have a way of collecting high quality, community-scale data over longer time periods. To do this, a long-term engagement built on trust is critical. The system must provide direct benefit for the community data partner. Lastly, the community must have a degree of data ownership or ability to limit who the data is shared with. If these requirements are met, there is the possibility for large institutions or companies building AI at scale to partner with communities.

4.7 Conclusion

In this project, we partnered with climate scientists and a Chicago community organization to develop a participatory AI system for precision climate. Through participant observation, interviews, and co-design workshops ("Community Cafes") we learned about the needs of each stakeholder. We ultimately co-designed and developed an app called *Water On My Block* to allow neighborhood residents to report flooding incidents that are both displayed on a website and are fed into the science AI models. We describe stakeholder needs for precision climate, challenges to bridging the gap between scientists and community members, and lessons learned in the development of the app.

CHAPTER 5

CONCLUSION

This dissertation has explored how AI is integrated into science and healthcare through a sociotechnical and human-centered lens. By examining three case studies—AI in home-based healthcare, AI adoption in a scientific research organization, and participatory AI for climate adaptation, I have demonstrated how AI's effectiveness is deeply intertwined with social contexts and organizational structures. Each case study highlights the complexity of designing AI that is not only technically robust but also socially attuned and inclusive.

The first study on AI-powered voice assistants for Black older adults underscored how users' relationships with AI systems influence their trust and engagement with healthcare interventions. While some participants embraced AI as a motivating health companion, others viewed it with skepticism, reinforcing the importance of designing AI that respects cultural and social dynamics. The second study on generative AI in a national lab revealed both the opportunities and risks of AI in scientific workplaces. Professionals across science and operations roles experimented with AI as a copilot, but concerns around security, trust, and institutional values shaped its adoption. Finally, the third study on participatory AI for climate adaptation illustrated the challenges and benefits of co-designing AI systems with historically marginalized communities. This project highlighted the need for long-term relationship-building, community ownership of data, and alignment between scientific and local knowledge.

Taken together, these studies contribute to the broader discourse on AI as a sociotechnical system. They demonstrate that AI is not a neutral tool but a social machine embedded in human relationships, organizational contexts, and historical inequalities. This work also reinforces the importance of inclusive design, ensuring that AI systems benefit those who are often left out of mainstream technological development.

5.1 Lessons Learned

Across the three case studies in this thesis, participants were primarily focused on solving familiar, persistent problems—such as staying healthy, managing information at work, or coping with urban flooding. While AI brought novel technical capabilities to these issues, participants cared less about the presence of AI itself and more about whether it tangibly helped improve their lives. Exercise, office work, and flooding are not new problems; what matters is whether a new approach offers real progress.

However, AI differs from prior tools: it often has opaque functionality, anthropomorphic interfaces, and unclear data flows. This creates unique challenges for everyday adoption. When people are offered help with serious, ongoing challenges—limited mobility, overwhelming workloads, or chronic infrastructure issues—they are generally open to trying new solutions. But when the technology starts behaving unpredictably (e.g. bringing up irrelevant topics like golf), threatens to replicate or replace their labor, or extracts data without obvious community benefit, trust becomes fragile. People are not only evaluating whether the tool works; they are assessing how it fits into their understanding of the social world, and how it might shift their own position within this world. Some participants related to AI as a helpful companion or co-worker, while others saw it as diminishing their value in the eyes of their colleagues or redistributing resources away from their communities.

From this, two lessons stand out: first, AI systems must provide clear, meaningful value that is legible and aligned with users' real needs. Second, AI system designers must recognize that they are intervening in social structures—potentially altering power dynamics based on decisions such as whose culture the AI mimics, what labor the AI can replicate, and where the AI system concentrates the benefit of collected data. Interface and interaction design choices can help guide users in understanding the purpose of an AI system, building appropriate expectations, and shaping how they relate to it.

Engaging with people in real-world environments also requires deep, sustained relationship-building. Trust does not come from design principles alone—it is built through time, presence, and accountability. This work involved significant in-person effort: conducting focus groups with older adults at a senior clinic, attending weekend climate events in Chatham, and coordinating interviews around the availability of clinical trial field teams. Reaching and respecting stakeholders required flexibility, patience, and a willingness to meet people on their own terms. Moreover, working across disciplines and community contexts required careful translation of goals and priorities to support shared understanding. I learned how to align diverse stakeholder needs and move toward a collective vision, and realized how essential this is for building inclusive, human-centered AI systems.

5.2 Future Work

This thesis leads to several avenues for future research on AI in social settings. As AI continues to integrate into daily life, workplaces, and community contexts, it is crucial to refine the ways these systems interact with people these social structures. Future work can build on the findings of this dissertation by deepening our understanding of AI's role in shaping interpersonal relationships, supporting organizational knowledge work, and fostering equitable data partnerships with communities.

5.2.1 Designing Culturally-Sensitive Interpersonal Relations with AI

As AI agents become increasingly embedded in people's daily routines, it will be essential to design these interactions with cultural and interpersonal sensitivities in mind. In Chapter 2, I explored how Black older adults related to an Amazon Alexa in various ways—some perceiving it as a friendly companion, while others viewed it as a stranger or even an intruder. These perceptions directly influenced their willingness to accept health-related advice from the AI. Future research should examine how diverse social groups relate to non-human entities and how AI systems can be designed to foster safe, contextually appropriate, and beneficial interpersonal relationships. Additionally, as AI begins to mediate human-to-human interactions—whether in the workplace or personal settings—it is important to understand how these agent-driven relationships affect trust, social bonds, and power dynamics. This is particularly critical in high-stakes environments like healthcare, where trust in AI could directly impact medical decision-making and patient outcomes.

5.2.2 Copilots for Organization Information Retrieval

AI copilots are poised to transform workplace knowledge management, but their design presents numerous HCI challenges. In Chapter 3, I found that one of the most sought-after applications for generative AI in a science organization was its ability to extract and summarize insights from vast amounts of unstructured text data. However, designing an effective AI system for organizational information retrieval involves addressing several key challenges: ensuring that the right information reaches the right user at the right time, integrating retrieval mechanisms into workflows in a seamless way, and preserving necessary contextual information to maintain relevance. Additionally, security concerns in science organizations—such as access controls for classified or proprietary information—raise further design considerations. Future research should investigate how AI can serve as a trustworthy and effective collaborator in information-rich work environments while balancing usability, security, and efficiency.

5.2.3 Community Data Partners

Existing AI systems often extract data from individuals—such as through crowdsourcing platforms like Amazon Mechanical Turk—without creating reciprocal value for communities. In Chapter 4, I examined a participatory AI system where community-collected data was not only used for AI model training but also made actionable for the community itself. Expanding on this work, future research could explore sustainable strategies for building long-term, mutually beneficial data partnerships with communities. Key challenges include structuring incentives for participation, ensuring that communities retain agency over their data, and designing AI systems that provide meaningful, tangible benefits in return. As AI applications shift from serving singular users to engaging with entire groups, designing ethical, community-centered approaches to data collection will be increasingly important.

5.3 Conclusion

This thesis has explored the integration of AI into healthcare and science through a sociotechnical and human-centered lens, demonstrating how these technologies interact with real-world social structures, organizational dynamics, and community needs. In healthcare, this work highlights the importance of trust, cultural sensitivity, and human-AI collaboration in designing systems that support well-being. In scientific research, it underscores how AI can enhance knowledge work while also introducing new risks and challenges in organizational contexts. By centering inclusion and participatory design, this work illustrates how AI systems can be both effectively and ethically integrated into these high-stakes domains. Given that AI continues to advance at an unprecedented pace, researchers, designers, and policymakers must work together to shape AI systems that are just, equitable, and attuned to the diverse ways humans live, work, and interact socially. This is particularly true in fields like science and healthcare, where the stakes are high and the potential for positive impact is large.

REFERENCES

[1] The world cafe method, 2015. URL https://theworldcafe.com/key-concepts-resources/world-cafe-method/.

[2] Chatham community data snapshot. Technical report, Chicago Metropolitan Agency for Planning, 2024.

[3] Generative ai in real-world workplaces: The second microsoft report on ai and productivity research. *Microsoft*, July 2024. URL https://www.microsoft.com/en-us/research/publication/generative-ai-in-real-world-workplaces/.

[4] M. S. Ackerman. The intellectual challenge of cscw: The gap between social requirements and technical feasibility. *Human–Computer Interaction*, 15(2–3):179–203, Sept. 2000. ISSN 0737-0024. doi:10.1207/S15327051HCI1523_5.

[5] E. Agapie, L. Colusso, S. A. Munson, and G. Hsieh. Plansourcing: Generating behavior change plans with friends and crowds. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, CSCW '16, page 119–133, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450335928. doi:10.1145/2818048.2819943. URL https://doi.org/10.1145/2818048.2819943.

[6] E. Agapie, B. Chinh, L. R. Pina, D. Oviedo, M. C. Welsh, G. Hsieh, and S. Munson. Crowdsourcing exercise plans aligned with expert guidelines and everyday constraints. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, page 1–13, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450356206. doi:10.1145/3173574.3173898. URL https://doi.org/10.1145/3173574.3173898.

[7] I. Ahmad, T. Akter, Z. Buher, R. Farzan, A. Kapadia, and A. J. Lee. Tangible privacy for smart voice assistants: Bystanders' perceptions of physical device controls. *Proc. ACM Hum.-Comput. Interact.*, 6(CSCW2), Nov. 2022. doi:10.1145/3555089. URL https://doi.org/10.1145/3555089.

[8] M. R. AI4Science and M. A. Quantum. The impact of large language models on scientific discovery: a preliminary study using gpt-4. (arXiv:2311.07361), November 2024. doi:10.48550/arXiv.2311.07361. URL http://arxiv.org/abs/2311.07361.

[9] S. Amershi, D. Weld, M. Vorvoreanu, A. Fourney, B. Nushi, P. Collisson, J. Suh, S. Iqbal, P. N. Bennett, K. Inkpen, J. Teevan, R. Kikin-Gil, and E. Horvitz. Guidelines for human-ai interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, page 1–13, New York, NY, USA, May 2019. Association for Computing Machinery. ISBN 978-1-4503-5970-2. doi:10.1145/3290605.3300233. URL https://dl.acm.org/doi/10.1145/3290605.3300233.

[10] E. Arias, B. Tejada-Vera, K. D. Kochanek, and F. B. Ahmad. Provisional Life Expectancy Estimates for 2021. Technical Report 23, National Vital Statistics System, Aug. 2022.

[11] M. Aristeidou, E. Scanlon, and M. Sharples. Weather-it: evolution of an online community for citizen inquiry. In *Proceedings of the 15th International Conference on Knowledge Technologies and Data-driven Business*, i-KNOW '15, page 1–8, New York, NY, USA, Oct 2015. Association for Computing Machinery. ISBN 978-1-4503-3721-2. doi:10.1145/2809563.2809567. URL `https://dl.acm.org/doi/10.1145/2809563.2809567`.

[12] A. Arnold, S. Kolody, A. Comeau, and A. Miguel Cruz. What does the literature say about the use of personal voice assistants in older adults? a scoping review. *Disability and Rehabilitation: Assistive Technology*, 19(1):100–111, Jan. 2024. ISSN 1748-3107. doi:10.1080/17483107.2022.2065369.

[13] J. Bardzell and S. Bardzell. Humanistic hci. *Interactions*, 23(2):20–29, Feb. 2016. doi:https://doi.org/10.1145/2888576. PUB27 New York, NY, USA.

[14] S. Bardzell. Utopias of participation: Feminism, design, and the futures. *ACM Transactions on Computer-Human Interaction*, 25(1):6:1–6:24, Feb. 2018. ISSN 1073-0516. doi:10.1145/3127359.

[15] S. Bardzell and J. Bardzell. Towards a feminist hci methodology: social science, feminism, and hci. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, page 675–684, Vancouver, BC, Canada, May 2011. Association for Computing Machinery. ISBN 978-1-4503-0228-9. doi:10.1145/1978942.1979041. URL `https://doi.org/10.1145/1978942.1979041`.

[16] N. K. Baym. *Personal Connections in the Digital Age*. Polity, Malden, MA, 1 edition edition, June 2010. ISBN 978-0-7456-4332-8.

[17] E. Beede, E. Baylor, F. Hersch, A. Iurchenko, L. Wilcox, P. Ruamviboonsuk, and L. M. Vardoulakis. A human-centered evaluation of a deep learning system deployed in clinics for the detection of diabetic retinopathy. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–12, New York, NY, USA, Apr. 2020. Association for Computing Machinery. ISBN 978-1-4503-6708-0. doi:10.1145/3313831.3376718. URL `https://doi.org/10.1145/3313831.3376718`.

[18] R. Benjamin. *Race After Technology: Abolitionist Tools for the New Jim Code*. Polity, Medford, MA, 2019. ISBN 978-1-5095-2640-6.

[19] T. W. Bickmore, L. Caruso, and K. Clough-Gorr. Acceptance and usability of a relational agent interface by urban older adults. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '05, page 1212–1215, New York, NY, USA, Apr. 2005. Association for Computing Machinery. ISBN 978-1-59593-002-6.

doi:10.1145/1056808.1056879. URL https://dl.acm.org/doi/10.1145/1056808.1056879.

[20] T. W. Bickmore, L. M. Pfeifer, and B. W. Jack. Taking the time to care: empowering low health literacy hospital patients with virtual nurse agents. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, page 1265–1274, New York, NY, USA, Apr 2009. Association for Computing Machinery. ISBN 978-1-60558-246-7. doi:10.1145/1518701.1518891. URL https://doi.org/10.1145/1518701.1518891.

[21] A. Birhane, W. Isaac, V. Prabhakaran, M. Díaz, M. C. Elish, I. Gabriel, and S. Mohamed. Power to the people? opportunities and challenges for participatory ai. Sep 2022. doi:10.1145/3551624.3555290. URL http://arxiv.org/abs/2209.07572. arXiv:2209.07572 [cs].

[22] K. Bonilla and A. Martin-Hammond. Older adults' perceptions of intelligent voice assistant privacy, transparency, and online privacy guidelines. In *Sixteenth Symposium on Usable Privacy and Security (SOUPS 2020)*, 2020.

[23] K. Bonilla, B. Gaitan, J. Sanders, N. Khenglawt, and A. Martin-Hammond. Comparing older and younger adults perceptions of voice and text-based search for consumer health information tasks. *AMIA Annual Symposium Proceedings*, 2021:227–236, Feb. 2022. ISSN 1942-597X.

[24] W. Boot. Benefits and barriers to emerging technologies to promote health, well-being, and independence of older adults. *Innovation in Aging*, 5(Suppl 1):265, Dec 2021. ISSN 2399-5300. doi:10.1093/geroni/igab046.1022.

[25] A. Borunda. Trump officials signal potential changes at noaa, the weather and climate agency. *NPR*, Feb. 2025. URL https://www.npr.org/2025/02/15/nx-s1-5297183/noaa-national-weather-service-trump.

[26] C. Bosco, F. Shojaei, A. A. Theisz, J. O. Torres, B. Cureton, A. K. Himes, N. M. Jessup, P. A. Barnes, Y. Lu, H. C. Hendrie, C. V. Hill, and P. C. Shih. Testing 3 modalities (voice assistant, chatbot, and mobile app) to assist older african american and black adults in seeking information on alzheimer disease and related dementias: Wizard of oz usability study. *JMIR Formative Research*, 8(1):e60650, Dec. 2024. doi:10.2196/60650.

[27] A. Bowser, K. Shilton, J. Preece, and E. Warrick. Accounting for privacy in citizen science: Ethical research in a context of openness. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '17, page 2124–2136, New York, NY, USA, Feb 2017. Association for Computing Machinery. ISBN 978-1-4503-4335-0. doi:10.1145/2998181.2998305. URL https://dl.acm.org/doi/10.1145/2998181.2998305.

[28] C. Bremer, B. Knowles, and A. Friday. Have we taken on too much?: A critical review of the sustainable hci landscape. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, page 1–11, New York, NY, USA, Apr 2022. Association for Computing Machinery. ISBN 978-1-4503-9157-3. doi:10.1145/3491102.3517609. URL `https://doi.org/10.1145/3491102.3517609`.

[29] R. Brewer, C. Pierce, P. Upadhyay, and L. Park. An empirical study of older adult's voice assistant use for health information seeking. *ACM Trans. Interact. Intell. Syst.*, 12(2), jul 2022. ISSN 2160-6455. doi:10.1145/3484507. URL `https://doi-org.proxy.uchicago.edu/10.1145/3484507`.

[30] R. N. Brewer. "if alexa knew the state i was in, it would cry": Older adults' perspectives of voice assistants for health. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI EA '22, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450391566. doi:10.1145/3491101.3519642. URL `https://doi.org/10.1145/3491101.3519642`.

[31] R. N. Brewer, C. Harrington, and C. Heldreth. Envisioning equitable speech technologies for black older adults. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '23, page 379–388, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9798400701924. doi:10.1145/3593013.3594005. URL `https://doi.org/10.1145/3593013.3594005`.

[32] S. Browne. *Dark Matters: On the Surveillance of Blackness*. Duke University Press, Durham, NC, 2015. ISBN 978-0-8223-5938-8.

[33] E. Brynjolfsson, D. Li, and L. R. Raymond. Generative ai at work. Working Paper 31161, National Bureau of Economic Research, Apr. 2023. URL `https://www.nber.org/papers/w31161`.

[34] J. Buolamwini and T. Gebru. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency*, page 77–91. PMLR, 2018.

[35] U. C. Bureau. Working-age population not keeping pace with growth in older americans, 2020. URL `https://www.census.gov/library/stories/2020/06/working-age-population-not-keeping-pace-with-growth-in-older-americans.html`.

[36] U. C. Bureau. 65 and older population grows rapidly as baby boomers age, 2020. URL `https://www.census.gov/newsroom/press-releases/2020/65-older-population-grows.html`.

[37] J. Butler, S. Jaffe, N. Baym, M. Czerwinski, S. Iqbal, K. Nowak, S. Rintel, A. Sellen, M. Vorvoreanu, B. Hecht, and J. Teevan. *Microsoft New Future of Work Report 2023*. 2023. URL `https://aka.ms/nfw2023`.

[38] C. J. Cai, E. Reif, N. Hegde, J. Hipp, B. Kim, D. Smilkov, M. Wattenberg, F. Viegas, G. S. Corrado, M. C. Stumpe, and M. Terry. 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*, CHI '19, page 1–14, New York, NY, USA, May 2019. Association for Computing Machinery. ISBN 978-1-4503-5970-2. doi:10.1145/3290605.3300234. URL `https://doi.org/10.1145/3290605.3300234`.

[39] C. J. Cai, S. Winter, D. Steiner, L. Wilcox, and M. Terry. "hello ai": Uncovering the onboarding needs of medical practitioners for human-ai collaborative decision-making. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):104:1–104:24, Nov. 2019. doi:10.1145/3359206.

[40] C. Caldeira, N. Nurain, A. A. Heintzman, H. Molchan, K. Caine, G. Demiris, K. A. Siek, B. Reeder, and K. Connelly. "how do i compare to the other people?": Older adults' perspectives on personal smart home data for self-management. *Proc. ACM Hum.-Comput. Interact.*, 7(CSCW2):238:1–238:32, Oct. 2023. doi:10.1145/3610029.

[41] T. Capel and M. Brereton. What is human-centered about human-centered ai? a map of the research landscape. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, page 1–23, New York, NY, USA, Apr. 2023. Association for Computing Machinery. ISBN 978-1-4503-9421-5. doi:10.1145/3544548.3580959. URL `https://dl.acm.org/doi/10.1145/3544548.3580959`.

[42] C. Carras. Major record labels sue ai companies for allegedly stealing copyrighted music. *Los Angeles Times*, June 2024. URL `https://www.latimes.com/entertainment-arts/business/story/2024-06-24/riaa-suno-udio-lawsuit-ai-copyright-songs-music`.

[43] L. A. Celi, J. Cellini, M.-L. Charpignon, E. C. Dee, F. Dernoncourt, R. Eber, W. G. Mitchell, L. Moukheiber, J. Schirmer, J. Situ, J. Paguio, J. Park, J. G. Wawira, S. Yao, and f. M. C. Data. Sources of bias in artificial intelligence that perpetuate healthcare disparities—a global review. *PLOS Digital Health*, 1(3):e0000022, Mar. 2022. ISSN 2767-3170. doi:10.1371/journal.pdig.0000022.

[44] C. Chen, J. G. Johnson, K. Charles, A. Lee, E. T. Lifset, M. Hogarth, A. A. Moore, E. Farcas, and N. Weibel. Understanding Barriers and Design Opportunities to Improve Healthcare and QOL for Older Adults through Voice Assistants. In *The 23rd International ACM SIGACCESS Conference on Computers and Accessibility*, pages 1–16, Virtual Event USA, Oct. 2021. ACM. ISBN 978-1-4503-8306-6. doi:10.1145/3441852.3471218. URL `https://dl.acm.org/doi/10.1145/3441852.3471218`.

[45] J. Chung, G. Demiris, H. J. Thompson, K.-Y. Chen, R. Burr, S. Patel, and J. Fogarty. Feasibility testing of a home-based sensor system to monitor mobility and daily activ-

ities in korean american older adults. *International Journal of Older People Nursing*, 12(1):e12127, 2017. doi:10.1111/opn.12127.

[46] D. E. Conroy, C.-H. Yang, and J. P. Maher. Behavior change techniques in top-ranked mobile apps for physical activity. *American journal of preventive medicine*, 46(6): 649–652, 2014. doi:10.1016/j.amepre.2014.01.010.

[47] S. Consolvo, K. Everitt, I. Smith, and J. A. Landay. Design requirements for technologies that encourage physical activity. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, page 457–466, New York, NY, USA, 2006. Association for Computing Machinery. ISBN 1595933727. doi:10.1145/1124772.1124840. URL `https://doi.org/10.1145/1124772.1124840`.

[48] K. Cortiñas-Lorenzo, S. Lindley, I. Larsen-Ledet, and B. Mitra. Through the looking-glass: Transparency implications and challenges in enterprise ai knowledge systems. (arXiv:2401.09410), Jan. 2024. doi:10.48550/arXiv.2401.09410. URL `http://arxiv.org/abs/2401.09410`.

[49] S. Costanza-Chock. *Design Justice: Community-Led Practices to Build the Worlds We Need*. The MIT Press, Cambridge, MA, Mar. 2020. ISBN 978-0-262-04345-8.

[50] M. Coutinho, L. Marques, A. Santos, M. Dahia, C. França, and R. de Souza Santos. The role of generative ai in software development productivity: A pilot case study. In *Proceedings of the 1st ACM International Conference on AI-Powered Software*, AIware 2024, page 131–138, New York, NY, USA, July 2024. Association for Computing Machinery. ISBN 9798400706851. doi:10.1145/3664646.3664773. URL `https://doi.org/10.1145/3664646.3664773`.

[51] K. Crawford. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press, New Haven, Apr. 2021. ISBN 978-0-300-20957-0.

[52] K. Crawford and V. Joler. *Anatomy of an AI System: The Amazon Echo As An Anatomical Map of Human Labor, Data and Planetary Resources*. Sept. 2018. URL `http://www.anatomyof.ai`.

[53] A. Crosby, E. C. Orenstein, S. E. Poulton, K. L. Bell, B. Woodward, H. Ruhl, K. Katija, and A. G. Forbes. Designing ocean vision ai: An investigation of community needs for imaging-based ocean conservation. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, page 1–16, New York, NY, USA, Apr. 2023. Association for Computing Machinery. ISBN 978-1-4503-9421-5. doi:10.1145/3544548.3580886. URL `https://dl.acm.org/doi/10.1145/3544548.3580886`.

[54] A. Cuadra, J. Bethune, R. Krell, A. Lempel, K. Hänsel, A. Shahrokni, D. Estrin, and N. Dell. Designing voice-first ambient interfaces to support aging in place. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*, DIS '23, page

2189–2205, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9781450398930. doi:10.1145/3563657.3596104. URL https://doi.org/10.1145/3563657.3596104.

[55] M. I. G. Daepp, A. Cabral, T. M. Werner, R. Mansour, C. Catlett, A. Roseway, C. Needham, N. Udeagbala, and S. Counts. The "three-legged stool": Designing for equitable city, community, and research partnerships in urban environmental sensing. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, page 1–19, New York, NY, USA, Apr 2023. Association for Computing Machinery. ISBN 978-1-4503-9421-5. doi:10.1145/3544548.3581289. URL https://dl.acm.org/doi/10.1145/3544548.3581289.

[56] M. T. Dearing, Y. Tao, X. Wu, Z. Lan, and V. Taylor. Lassi: An llm-based automated self-correcting pipeline for translating parallel scientific codes. (arXiv:2407.01638), June 2024. doi:10.48550/arXiv.2407.01638. URL http://arxiv.org/abs/2407.01638.

[57] F. Delgado, S. Yang, M. Madaio, and Q. Yang. Stakeholder participation in ai: Beyond "add diverse stakeholders and stir". (arXiv:2111.01122), Nov 2021. doi:10.48550/arXiv.2111.01122. URL http://arxiv.org/abs/2111.01122. arXiv:2111.01122 [cs].

[58] F. Delgado, S. Yang, M. Madaio, and Q. Yang. The participatory turn in AI design: Theoretical foundations and the current state of practice. In *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, EAAMO '23, page 1–23, New York, NY, USA, Oct. 2023. Association for Computing Machinery. ISBN 9798400703812. doi:10.1145/3617694.3623261. URL https://dl.acm.org/doi/10.1145/3617694.3623261.

[59] F. Dell'Acqua, E. McFowland III, E. R. Mollick, H. Lifshitz-Assaf, K. Kellogg, S. Rajendran, L. Krayer, F. Candelon, and K. R. Lakhani. Navigating the jagged technological frontier: Field experimental evidence of the effects of ai on knowledge worker productivity and quality. Working Paper 24-013, Harvard Business School Technology and Operations Management, Sept. 2023. URL https://papers.ssrn.com/abstract=4573321.

[60] G. Demiris, M. J. Rantz, M. A. Aud, K. D. Marek, H. W. Tyrer, M. Skubic, and A. A. Hussam. Older adults' attitudes towards and perceptions of 'smart home' technologies: A pilot study. *Medical Informatics and the Internet in Medicine*, 29(2):87–94, June 2004. ISSN 1463-9238. doi:10.1080/14639230410001684387.

[61] M. A. DeVito. Adaptive folk theorization as a path to algorithmic literacy on changing platforms. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2): 339:1–339:38, Oct 2021. doi:10.1145/3476080.

[62] Z. Diao, H. Yamashita, and M. Abe. Leveraging large language models and social media for automation in scanning probe microscopy. (arXiv:2405.15490), 2024. doi:10.48550/arXiv.2405.15490. URL https://arxiv.org/abs/2405.15490.

[63] J. Dinet and M. Kitajima. “draw me the web”: impact of mental model of the web on information search performance of young users. In *Proceedings of the 23rd Conference on l’Interaction Homme-Machine*, IHM ’11, page 1–7, New York, NY, USA, Oct 2011. Association for Computing Machinery. ISBN 978-1-4503-0822-9. doi:10.1145/2044354.2044358. URL https://doi.org/10.1145/2044354.2044358.

[64] C. DiSalvo. *Adversarial Design*. The MIT Press, Cambridge; London, reprint edition edition, Aug. 2015. ISBN 978-0-262-52822-1.

[65] P. Dourish. *Where the Action Is: The Foundations of Embodied Interaction*. MIT Press, 2004. URL https://www.amazon.com/Where-Action-MIT-Press-Foundations/dp/0262541785.

[66] I. Drosos, A. Sarkar, X. Xu, C. Negreanu, S. Rintel, and L. Tankelevitch. “it’s like a rubber duck that talks back”: Understanding generative ai-assisted data analysis workflows through a participatory prompting study. In *Proceedings of the 3rd Annual Meeting of the Symposium on Human-Computer Interaction for Work*, CHIWORK ’24, page 1–21, New York, NY, USA, June 2024. Association for Computing Machinery. ISBN 9798400710179. doi:10.1145/3663384.3663389. URL https://dl.acm.org/doi/10.1145/3663384.3663389.

[67] A. Dunne and F. Raby. *Speculative Everything: Design, Fiction, and Social Dreaming*. The MIT Press, Cambridge, Massachusetts; London, illustrated edition edition, Dec. 2013. ISBN 978-0-262-01984-2.

[68] Y. K. Dwivedi, N. Kshetri, L. Hughes, E. L. Slade, A. Jeyaraj, A. K. Kar, A. M. Baabdullah, A. Koohang, V. Raghavan, M. Ahuja, H. Albanna, M. A. Albashrawi, A. S. Al-Busaidi, J. Balakrishnan, Y. Barlette, S. Basu, I. Bose, L. Brooks, D. Buhalis, L. Carter, S. Chowdhury, T. Crick, S. W. Cunningham, G. H. Davies, R. M. Davison, R. Dé, D. Dennehy, Y. Duan, R. Dubey, R. Dwivedi, J. S. Edwards, C. Flavián, R. Gauld, V. Grover, M.-C. Hu, M. Janssen, P. Jones, I. Junglas, S. Khorana, S. Kraus, K. R. Larsen, P. Latreille, S. Laumer, F. T. Malik, A. Mardani, M. Mariani, S. Mithas, E. Mogaji, J. H. Nord, S. O’Connor, F. Okumus, M. Pagani, N. Pandey, S. Papagiannidis, I. O. Pappas, N. Pathak, J. Pries-Heje, R. Raman, N. P. Rana, S.-V. Rehm, S. Ribeiro-Navarrete, A. Richter, F. Rowe, S. Sarker, B. C. Stahl, M. K. Tiwari, W. van der Aalst, V. Venkatesh, G. Viglia, M. Wade, P. Walton, J. Wirtz, and R. Wright. Opinion paper: “so what if chatgpt wrote it?” multidisciplinary perspectives on opportunities, challenges and implications of generative conversational ai for research, practice and policy. *International Journal of Information Management*, 71(C), Aug. 2023. ISSN 0268-4012. doi:10.1016/j.ijinfomgt.2023.102642. URL https://doi.org/10.1016/j.ijinfomgt.2023.102642.

[69] C. D'Ignazio and L. F. Klein. *Data Feminism*. The MIT Press, Feb. 2020.

[70] P. Ehn. Work-oriented design of computer artifacts. 1988. URL `https://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-62913`.

[71] S. L. Erete. Community, group and individual: A framework for designing community technologies. *The Journal of Community Informatics*, 10(11), 2014. ISSN 1712-4441. doi:10.15353/joci.v10i1.2671. URL `https://openjournals.uwaterloo.ca/index.php/JoCI/article/view/2671`.

[72] V. Eubanks. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin's Press, New York, NY, Jan. 2018. ISBN 978-1-250-07431-7.

[73] R. Fogliato, S. Chappidi, M. Lungren, P. Fisher, D. Wilson, M. Fitzke, M. Parkinson, E. Horvitz, K. Inkpen, and B. Nushi. Who goes first? influences of human-ai workflow on decision making in clinical imaging. In *2022 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '22, page 1362–1374, New York, NY, USA, June 2022. Association for Computing Machinery. ISBN 978-1-4503-9352-2. doi:10.1145/3531146.3533193. URL `https://doi.org/10.1145/3531146.3533193`.

[74] B. Friedman. Value-sensitive design. *Interactions*, 3(6):16–23, 1996.

[75] B. Friedman and D. G. Hendry. *Value Sensitive Design: Shaping Technology with Moral Imagination*. The MIT Press, Cambridge, Massachusetts, illustrated edition edition, May 2019. ISBN 978-0-262-03953-6.

[76] A. Frik, J. Bernd, N. Alomar, and S. Egelman. A qualitative model of older adults' contextual decision-making about information sharing. In *Workshop on the Economics of Information Security (WEIS 2020)*, 2020.

[77] A. Frik, J. Bernd, and S. Egelman. A Model of Contextual Factors Affecting Older Adults' Information-Sharing Decisions in the US. *ACM Transactions on Computer-Human Interaction*, page 3557888, Aug. 2022. ISSN 1073-0516, 1557-7325. doi:10.1145/3557888.

[78] J. Gabrys. *Citizens of Worlds: Open-Air Toolkits for Environmental Struggle*. Univ Of Minnesota Press, Minneapolis, Nov 2022. ISBN 978-1-5179-1405-9.

[79] Y. K. Gal, A. Segal, E. Kamar, E. Horvitz, C. Lintott, and M. Walmsley. A new workflow for human-ai collaboration in citizen science. In *Proceedings of the 2022 ACM Conference on Information Technology for Social Good*, GoodIT '22, page 89–95, New York, NY, USA, Sep 2022. Association for Computing Machinery. ISBN 978-1-4503-9284-6. doi:10.1145/3524458.3547243. URL `https://doi.org/10.1145/3524458.3547243`.

[80] Y. Gao, Z. Pan, H. Wang, and G. Chen. Alexa, my love: Analyzing reviews of amazon echo. In *2018 IEEE SmartWorld, Ubiquitous Intelligence Computing, Advanced Trusted Computing, Scalable Computing Communications, Cloud Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, page 372–380, Oct. 2018. doi:10.1109/SmartWorld.2018.00094. URL https://ieeexplore.ieee.org/abstract/document/8560072.

[81] K. Gero, A. Calderwood, C. Li, and L. Chilton. A design space for writing support tools using a cognitive process model of writing. In T.-H. uang, V. Raheja, D. Kang, J. J. Y. Chung, D. Gissin, M. Lee, and K. I. Gero, editors, *Proceedings of the First Workshop on Intelligent and Interactive Writing Assistants (In2Writing 2022)*, page 11–24, Dublin, Ireland, May 2022. Association for Computational Linguistics. doi:10.18653/v1/2022.in2writing-1.2. URL https://aclanthology.org/2022.in2writing-1.2.

[82] R. Ghaiumy Anaraky, K. A. Byrne, P. J. Wisniewski, X. Page, and B. Knijnenburg. To Disclose or Not to Disclose: Examining the Privacy Decision-Making Processes of Older vs. Younger Adults. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–14, Yokohama Japan, May 2021. ACM. ISBN 978-1-4503-8096-6. doi:10.1145/3411764.3445204.

[83] M. Glickman and Y. Zhang. Ai and generative ai for research discovery and summarization. *Harvard Data Science Review*, 6(2), Mar. 2024. doi:10.1162/99608f92.7f9220ff. URL https://hdsr.mitpress.mit.edu/pub/xedo5giw.

[84] W. Godoy, P. Valero-Lara, K. Teranishi, P. Balaprakash, and J. Vetter. Evaluation of openai codex for hpc parallel programming models kernel generation. In *Proceedings of the 52nd International Conference on Parallel Processing Workshops*, pages 136–144, Salt Lake Citym UT USA, August 2023. Association for Computing Machinery. doi:10.1145/3605731.3605886.

[85] G. Goth. Why are lawyers afraid of ai? *Commun. ACM*, 67(1):14–16, Dec. 2023. ISSN 0001-0782. doi:10.1145/3631935.

[86] M. L. Gray and S. Suri. *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*. Houghton Mifflin Harcourt, Boston, May 2019. ISBN 978-1-328-56624-9.

[87] C. N. Harrington and L. Egede. Trust, comfort and relatability: Understanding black older adults' perceptions of chatbot design for health information seeking. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, page 1–18, New York, NY, USA, Apr. 2023. Association for Computing Machinery. ISBN 978-1-4503-9421-5. doi:10.1145/3544548.3580719. URL https://dl.acm.org/doi/10.1145/3544548.3580719.

[88] C. N. Harrington, R. Garg, A. Woodward, and D. Williams. "it's kind of like code-switching": Black older adults' experiences with a voice assistant for health information seeking. page 15, 2022. doi:https://doi.org/10.1145/3491102.3501995.

[89] D. Harris, S. D. Dlima, A. Gluchowski, A. Hall, E. Elliott, and L. Munford. The effectiveness and acceptability of physical activity interventions amongst older adults with lower socioeconomic status: a mixed methods systematic review. *International Journal of Behavioral Nutrition and Physical Activity*, 21(1):121, Oct. 2024. ISSN 1479-5868. doi:10.1186/s12966-024-01666-8.

[90] M. Heikkilä. This new data poisoning tool lets artists fight back against generative ai. *MIT Technology Review*, 2023. URL `https://www.technologyreview.com/2023/10/23/1082189/data-poisoning-artists-fight-generative-ai/`.

[91] J. P. Higgins. Smartphone applications for patients' health and fitness. *The American journal of medicine*, 129(1):11–19, 2016. doi:10.1016/j.amjmed.2015.05.038.

[92] E. C. Ho, L. Hawkley, W. Dale, L. Waite, and M. Huisingh-Scheetz. Social capital predicts accelerometry-measured physical activity among older adults in the u.s.: a cross-sectional study in the national social life, health, and aging project. *BMC Public Health*, 18(1):804, Jun 2018. ISSN 1471-2458. doi:10.1186/s12889-018-5664-6.

[93] M. Hoff. *Draw What You Think Alexa Looks Like*. 2019. URL `https://docs.google.com/document/u/1/d/e/2PACX-1vS02YTBRsxnA5bmdtI5HdAwJ5raAzvcEe5cvWjIx_njRx9DvNt9rxJmZZ9BfvNT0eVvIdoguM6eLmxP/pub`.

[94] A. C. Horstmann, T. Schubert, L. Lambrich, and C. Strathmann. 'alexa, i do not want to be patronized': A qualitative interview study to explore older adults' attitudes towards intelligent voice assistants. In *Proceedings of the 23rd ACM International Conference on Intelligent Virtual Agents*, IVA '23, page 1–10, New York, NY, USA, Dec. 2023. Association for Computing Machinery. ISBN 978-1-4503-9994-4. doi:10.1145/3570945.3607342. URL `https://dl.acm.org/doi/10.1145/3570945.3607342`.

[95] X. Hu, S. Desai, M. Lundy, and J. Chin. Beyond functionality: Co-designing voice user interfaces for older adults' well-being, 2024. URL `https://arxiv.org/abs/2409.08449`.

[96] P. Hämäläinen, M. Tavast, and A. Kunnari. Evaluating large language models in generating synthetic hci research data: a case study. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, page 1–19, New York, NY, USA, Apr. 2023. Association for Computing Machinery. ISBN 978-1-4503-9421-5. doi:10.1145/3544548.3580688. URL `https://dl.acm.org/doi/10.1145/3544548.3580688`.

[97] A. Ismail and N. Kumar. Ai in global health: The view from the front lines. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*,

CHI '21, page 1–21, New York, NY, USA, May 2021. Association for Computing Machinery. ISBN 978-1-4503-8096-6. doi:10.1145/3411764.3445130. URL https://doi.org/10.1145/3411764.3445130.

[98] S. J. Jackson. Rethinking repair. In T. Gillespie, P. J. Boczkowski, and K. A. Foot, editors, *Media Technologies: Essays on Communication, Materiality, and Society*. MIT Press, Jan 2014. ISBN 978-0-262-31947-8.

[99] D. Jakob. Voice controlled devices and older adults–a systematic literature review. In *International Conference on Human-Computer Interaction*, pages 175–200. Springer, 2022. doi:https://doi.org/10.1007/978-3-031-05581-2_14.

[100] A. B. Jnr. Examining the use of intelligent conversational voice-assistants for improved mobility behavior of older adults in smart cities. 2024.

[101] J. Kaddour, J. Harris, M. Mozes, H. Bradley, R. Raileanu, and R. McHardy. Challenges and applications of large language models. (arXiv:2307.10169), July 2023. doi:10.48550/arXiv.2307.10169. URL http://arxiv.org/abs/2307.10169.

[102] M. Kazemitabaar, J. Chow, C. K. T. Ma, B. J. Ericson, D. Weintrop, and T. Grossman. Studying the effect of ai code generators on supporting novice learners in introductory programming. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, page 1–23, New York, NY, USA, Apr. 2023. Association for Computing Machinery. ISBN 978-1-4503-9421-5. doi:10.1145/3544548.3580919. URL https://dl.acm.org/doi/10.1145/3544548.3580919.

[103] D. Kim. Can healthcare apps and smart speakers improve the health behavior and depression of older adults? a quasi-experimental study. *Frontiers in Digital Health*, 5: 1117280, 2023. doi:https://doi.org/10.3389/fdgth.2023.1117280.

[104] J. Kim, S. Park, and L. Robert. Conversational agents for health and wellbeing: Review and future agendas. CSCW 2019, Oct 2019. URL http://deepblue.lib.umich.edu/handle/2027.42/151800. Accepted: 2019-10-29T09:59:51Z.

[105] J. Kim, S. Suh, L. B. Chilton, and H. Xia. Metaphorian: Leveraging large language models to support extended metaphor creation for science writing. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*, DIS '23, page 115–135, New York, NY, USA, July 2023. Association for Computing Machinery. ISBN 978-1-4503-9893-0. doi:10.1145/3563657.3595996. URL https://dl.acm.org/doi/10.1145/3563657.3595996.

[106] S. Kim. Exploring How Older Adults Use a Smart Speaker–Based Voice Assistant in Their First Interactions: Qualitative Study. *JMIR mHealth and uHealth*, 9(1):e20427, Jan. 2021. ISSN 2291-5222. doi:10.2196/20427. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7840274/.

[107] S. Kim and A. Choudhury. Exploring older adults' perception and use of smart speaker-based voice assistants: A longitudinal study. *Computers in Human Behavior*, 124: 106914, 2021. doi:https://doi.org/10.1016/j.chb.2021.106914.

[108] Y.-H. Kim, D. Chou, B. Lee, M. Danilovich, A. Lazar, D. E. Conroy, H. Kacorri, and E. K. Choe. Mymove: Facilitating older adults to collect in-situ activity labels on a smartwatch with speech. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, page 1–21, New York, NY, USA, Apr. 2022. Association for Computing Machinery. ISBN 978-1-4503-9157-3. doi:10.1145/3491102.3517457. URL `https://dl.acm.org/doi/10.1145/3491102.3517457`.

[109] B. Knowles and V. L. Hanson. Older Adults' Deployment of 'Distrust'. *ACM Transactions on Computer-Human Interaction*, 25(4):1–25, Aug. 2018. ISSN 1073-0516, 1557-7325. doi:10.1145/3196490.

[110] C. Kobiella, Y. S. Flores López, F. Waltenberger, F. Draxler, and A. Schmidt. "if the machine is as good as me, then what use am i?" – how the use of chatgpt changes young professionals' perception of productivity and accomplishment. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI '24, page 1–16, New York, NY, USA, May 2024. Association for Computing Machinery. ISBN 9798400703300. doi:10.1145/3613904.3641964. URL `https://dl.acm.org/doi/10.1145/3613904.3641964`.

[111] J. Kropczynski, Z. Aljallad, N. J. Elrod, H. Lipford, and P. J. Wisniewski. Towards Building Community Collective Efficacy for Managing Digital Privacy and Security within Older Adult Communities. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW3):1–27, Jan. 2021. ISSN 2573-0142. doi:10.1145/3432954.

[112] H. Langer and S. Merkel. Co-designing a conversational agent in health and social care–perspectives of professional care staff and older adults. In *International Conference on Human-Computer Interaction*, pages 191–202. Springer, 2023. doi:https://doi.org/10.1007/978-3-031-35921-7_13.

[113] L. Laranjo, A. G. Dunn, H. L. Tong, A. B. Kocaballi, J. Chen, R. Bashir, D. Surian, B. Gallego, F. Magrabi, A. Y. S. Lau, and E. Coiera. Conversational agents in healthcare: a systematic review. *Journal of the American Medical Informatics Association*, 25(9):1248–1258, Sept. 2018. ISSN 1067-5027, 1527-974X. doi:10.1093/jamia/ocy072. URL `https://academic.oup.com/jamia/article/25/9/1248/5052181`.

[114] M. Lee, K. I. Gero, J. J. Y. Chung, S. B. Shum, V. Raheja, H. Shen, S. Venugopalan, T. Wambsganss, D. Zhou, E. A. Alghamdi, T. August, A. Bhat, M. Z. Choksi, S. Dutta, J. L. Guo, M. N. Hoque, Y. Kim, S. Knight, S. P. Neshaei, A. Shibani, D. Shrivastava, L. Shroff, A. Sergeyuk, J. Stark, S. Sterman, S. Wang, A. Bosselut, D. Buschek, J. C. Chang, S. Chen, M. Kreminski, J. Park, R. Pea, E. H. R. Rho, Z. Shen, and P. Siangliulue. A design space for intelligent and interactive writing assistants. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI

'24, page 1–35, New York, NY, USA, May 2024. Association for Computing Machinery. ISBN 9798400703300. doi:10.1145/3613904.3642697. URL https://dl.acm.org/doi/10.1145/3613904.3642697.

[115] G. Lei, R. Docherty, and S. J. Cooper. Materials science in the era of large language models: a perspective. digital discovery. *Digital Discovery*, 3(7):1257–1272, 2024. doi:10.1039/D4DD00074A.

[116] S. Lohr. Generative a.i.'s biggest impact will be in banking and tech, report says. *The New York Times*, Feb. 2024. ISSN 0362-4331. URL https://www.nytimes.com/2024/02/01/business/ai-impact-jobs.html.

[117] A. Loten. Generative ai tools use custom data to power more business functions. *Wall Street Journal*, Mar. 2023. ISSN 0099-9660. URL https://www.wsj.com/articles/generative-ai-tools-use-custom-data-to-power-more-business-functions-2f158dbb.

[118] P. Maddigan and T. Susnjak. Chat2vis: Generating data visualizations via natural language using chatgpt, codex and gpt-3 large language models. *IEEE Access*, 11: 45181–45193, 2023. ISSN 2169-3536. doi:10.1109/ACCESS.2023.3274199.

[119] C. Marvin. *When Old Technologies Were New: Thinking About Electric Communication in the Late Nineteenth Century*. Oxford University Press, New York, NY, 1988. ISBN 978-0-19-506341-7.

[120] N. Mathur, K. Dhodapkar, T. Zubatiy, J. Li, B. Jones, and E. Mynatt. A collaborative approach to support medication management in older adults with mild cognitive impairment using conversational assistants (cas). In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '22, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450392587. doi:10.1145/3517428.3544830. URL https://doi.org/10.1145/3517428.3544830.

[121] N. McDonald, S. Schoenebeck, and A. Forte. Reliability and inter-rater reliability in qualitative research: Norms and guidelines for cscw and hci practice. *Proc. ACM Hum.-Comput. Interact.*, 3(CSCW), nov 2019. doi:10.1145/3359174. URL https://doi.org/10.1145/3359174.

[122] J. P. Migneault, J. J. Dedier, J. A. Wright, T. Heeren, M. K. Campbell, D. E. Morisky, P. Rudd, and R. H. Friedman. A culturally adapted telecommunication system to improve physical activity, diet quality, and medication adherence among hypertensive african–americans: A randomized controlled trial. *Annals of behavioral medicine: a publication of the Society of Behavioral Medicine*, 43(1):62–73, Feb. 2012. ISSN 0883-6612. doi:10.1007/s12160-011-9319-4.

[123] M. R. Morris. Scientists' perspectives on the potential for generative ai in their fields. (arXiv:2304.01420), Apr. 2023. doi:10.48550/arXiv.2304.01420. URL http://arxiv.org/abs/2304.01420. arXiv:2304.01420 [cs].

[124] J. E. Muñoz. *Cruising Utopia: The Then and There of Queer Futurity*. NYU Press, New York, Sep 2009. ISBN 978-0-8147-5728-4.

[125] P. Nallam, S. Bhandari, J. Sanders, and A. Martin-Hammond. A Question of Access: Exploring the Perceived Benefits and Barriers of Intelligent Voice Assistants for Improving Access to Consumer Health Resources Among Low-Income Older Adults. *Gerontology and Geriatric Medicine*, 6:233372142098597, Jan. 2020. ISSN 2333-7214, 2333-7214. doi:10.1177/2333721420985975.

[126] D. Nichols, J. H. Davis, Z. Xie, A. Rajaram, and A. Bhatele. Can large language models write parallel code? In *Proceedings of the 33rd International Symposium on High-Performance Parallel and Distributed Computing*, pages 281–294, Pisa, Italy, June 2024. Association for Computing Machinery. doi:10.1145/3625549.3658689.

[127] S. Noble. *Algorithms of Oppression: How Search Engines Reinforce Racism*. NYU Press, New York, 2018. ISBN 978-1-4798-3724-3.

[128] D. Norman. *The Design of Everyday Things*. Basic Books, 1990. URL `https://www.amazon.com/Design-Everyday-Things-Revised-Expanded/dp/0465050654`.

[129] S. Noy and W. Zhang. Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, 381(6654):187–192, July 2023. doi:10.1126/science.adh2586.

[130] N. Nurain and C.-F. Chung. "i left my legacy, told my story": Understanding older adults' tracking practices to promote active aging. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*, DIS '23, page 459–475, New York, NY, USA, July 2023. Association for Computing Machinery. ISBN 978-1-4503-9893-0. doi:10.1145/3563657.3596083. URL `https://dl.acm.org/doi/10.1145/3563657.3596083`.

[131] M. Odlum, N. Moise, I. M. Kronish, P. Broadwell, C. Alcántara, N. J. Davis, Y. K. K. Cheung, A. Perotte, and S. Yoon. Trends in Poor Health Indicators Among Black and Hispanic Middle-aged and Older Adults in the United States, 1999-2018. *JAMA Network Open*, 3(11):e2025134–e2025134, Nov. 2020. ISSN 2574-3805. doi:10.1001/jamanetworkopen.2020.25134. URL `https://doi.org/10.1001/jamanetworkopen.2020.25134`.

[132] B. Oewel, T. Ammari, and R. N. Brewer. Voice assistant use in long-term care. In *Proceedings of the 5th International Conference on Conversational User Interfaces*, CUI '23, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9798400700149. doi:10.1145/3571884.3597135. URL `https://doi.org/10.1145/3571884.3597135`.

[133] U. D. of Health and H. Services. Physical activity guidelines advisory committee scientific report, 2018.

[134] J. S. Park, J. O'Brien, C. J. Cai, M. R. Morris, P. Liang, and M. S. Bernstein. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, UIST '23, page 1–22, New York, NY, USA, Oct. 2023. Association for Computing Machinery. ISBN 9798400701320. doi:10.1145/3586183.3606763. URL https://dl.acm.org/doi/10.1145/3586183.3606763.

[135] S. Y. Park, P.-Y. Kuo, A. Barbarin, E. Kaziunas, A. Chow, K. Singh, L. Wilcox, and W. S. Lasecki. Identifying challenges and opportunities in human-ai collaboration in healthcare. In *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing*, CSCW '19, page 506–510, New York, NY, USA, Nov. 2019. Association for Computing Machinery. ISBN 978-1-4503-6692-2. doi:10.1145/3311957.3359433. URL https://doi.org/10.1145/3311957.3359433.

[136] S. Passi and M. Vorvoreanu. Overreliance on ai: Literature review. Microsoft technical report msr-tr-2022-12, Microsoft, June 2022. URL https://www.microsoft.com/en-us/research/publication/overreliance-on-ai-literature-review/.

[137] H. Pearce, B. Ahmad, B. Tan, B. Dolan-Gavitt, and R. Karri. Asleep at the keyboard? assessing the security of github copilot's code contributions. In *2022 IEEE Symposium on Security and Privacy (SP)*, page 754–768, May 2022. doi:10.1109/SP46214.2022.9833571. URL https://ieeexplore.ieee.org/abstract/document/9833571?casa_token=nu3RFriTDRMAAAAA:SVfp52xnsyyEJ9nweAEeGg6ysfXgqQoBy0QGhAJWpiSM4qFU3JkGrPZMNd8_I4sk-gk6zziXnA.

[138] S. Pednekar, P. Dhirawani, R. Shah, N. Shekokar, and K. Ghag. Voice-based interaction for an aging population: A systematic review. In *2023 3rd International Conference on Intelligent Communication and Computational Techniques (ICCT)*, page 1–8, Jan. 2023. doi:10.1109/ICCT56969.2023.10075801. URL https://ieeexplore.ieee.org/abstract/document/10075801#citations.

[139] L. Pei, M. Wong-Villacres, S. Erete, D. Rosner, A. Taylor, and M. Wiberg. Who we are and what we have: Designing with minoritized communities. *Interactions*, 29(5), 2022. URL https://interactions.acm.org/archive/view/september-october-2022/who-we-are-and-what-we-have.

[140] S. Peng, E. Kalliamvakou, P. Cihon, and M. Demirer. The impact of ai on developer productivity: Evidence from github copilot. (arXiv:2302.06590), Feb. 2023. doi:10.48550/arXiv.2302.06590. URL http://arxiv.org/abs/2302.06590.

[141] N. Perry, M. Srivastava, D. Kumar, and D. Boneh. Do users write more insecure code with ai assistants? In *Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security*, page 2785–2799, Nov. 2023. doi:10.1145/3576915.3623157. URL http://arxiv.org/abs/2211.03622.

[142] S. Petridis, N. Diakopoulos, K. Crowston, M. Hansen, K. Henderson, S. Jastrzebski, J. V. Nickerson, and L. B. Chilton. Anglekindling: Supporting journalistic angle ideation with large language models. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, page 1–16, New York, NY, USA, Apr. 2023. Association for Computing Machinery. ISBN 978-1-4503-9421-5. doi:10.1145/3544548.3580907. URL `https://dl.acm.org/doi/10.1145/3544548.3580907`.

[143] H. Petrie. Talking 'bout my generation ... or not? the digital technology life experiences of older people. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI EA '23, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9781450394222. doi:10.1145/3544549.3582742. URL `https://doi.org/10.1145/3544549.3582742`.

[144] T. Phan. Amazon echo and the aesthetics of whiteness. *Catalyst: Feminism, Theory, Technoscience*, 5(1):1–38, Apr. 2019. ISSN 2380-3312, 2380-3312. doi:10.28968/cftt.v5i1.29586.

[145] E. S. Poole, M. Chetty, R. E. Grinter, and W. K. Edwards. More than meets the eye: Transforming the user experience of home network management. In *Proceedings of the 7th ACM conference on Designing interactive systems*, DIS '08, page 455–464, New York, NY, USA, Feb 2008. Association for Computing Machinery. ISBN 978-1-60558-002-9. doi:10.1145/1394445.1394494. URL `https://doi.org/10.1145/1394445.1394494`.

[146] A. Pradhan, K. Mehta, and L. Findlater. *"Accessibility Came by Accident:" Use of Voice-Controlled Intelligent Personal Assistants by People with Disabilities*, page 1–13. Association for Computing Machinery, New York, NY, USA, Apr 2018. ISBN 978-1-4503-5620-6. URL `https://doi.org/10.1145/3173574.3174033`.

[147] A. Pradhan, L. Findlater, and A. Lazar. "phantom friend" or "just a box with information": Personification and ontological categorization of smart speaker-based voice assistants by older adults. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):214:1–214:21, Nov 2019. doi:10.1145/3359316.

[148] A. Pradhan, A. Lazar, and L. Findlater. Use of intelligent voice assistants by older adults with low technology use. *ACM Transactions on Computer-Human Interaction*, 27(4):31:1–31:27, Sep 2020. ISSN 1073-0516. doi:10.1145/3373759.

[149] N. R. Prestopnik and K. Crowston. Citizen science system assemblages: understanding the technologies that support crowdsourced science. In *Proceedings of the 2012 iConference*, iConference '12, page 168–176, New York, NY, USA, Feb 2012. Association for Computing Machinery. ISBN 978-1-4503-0782-6. doi:10.1145/2132176.2132198. URL `https://doi.org/10.1145/2132176.2132198`.

[150] M. H. Prince, H. Chan, A. Vriza, T. Zhou, V. K. Sastry, M. T. Dearing, R. J. Harder, R. K. Vasudevan, and M. J. Cherukara. Opportunities for retrieval and tool augmented large language models in scientific facilities. (arXiv:2312.01291), 2023. doi:10.48550/arXiv.2312.01291. URL https://arxiv.org/abs/2312.01291.

[151] D. Qaurooni, A. Ghazinejad, I. Kouper, and H. Ekbia. Citizens for science and science for citizens: The view from participatory design. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, page 1822–1826, New York, NY, USA, May 2016. Association for Computing Machinery. ISBN 978-1-4503-3362-7. doi:10.1145/2858036.2858575. URL https://dl.acm.org/doi/10.1145/2858036.2858575.

[152] O. O. Queerinai, A. Ovalle, A. Subramonian, A. Singh, C. Voelcker, D. J. Sutherland, D. Locatelli, E. Breznik, F. Klubicka, H. Yuan, H. J, H. Zhang, J. Shriram, K. Lehman, L. Soldaini, M. Sap, M. P. Deisenroth, M. L. Pacheco, M. Ryskina, M. Mundt, M. Agarwal, N. Mclean, P. Xu, A. Pranav, R. Korpan, R. Ray, S. Mathew, S. Arora, S. John, T. Anand, V. Agrawal, W. Agnew, Y. Long, Z. J. Wang, Z. Talat, A. Ghosh, N. Dennler, M. Noseworthy, S. Jha, E. Baylor, A. Joshi, N. Y. Bilenko, A. Mcnamara, R. Gontijo-Lopes, A. Markham, E. Dong, J. Kay, M. Saraswat, N. Vytla, and L. Stark. Queer in ai: A case study in community-led participatory ai. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '23, page 1882–1895, New York, NY, USA, June 2023. Association for Computing Machinery. ISBN 9798400701924. doi:10.1145/3593013.3594134. URL https://dl.acm.org/doi/10.1145/3593013.3594134.

[153] A. R. Quiñones, G. J. McAvay, K. D. Peak, B. Vander Wyk, and H. G. Allore. The Contribution of Chronic Conditions to Hospitalization, Skilled Nursing Facility Admission, and Death: Variation by Race. *American Journal of Epidemiology*, 191(12): 2014–2025, Nov. 2022. ISSN 0002-9262. doi:10.1093/aje/kwac143.

[154] P. Rao, A. Joshi, and N. D'Souza. Investigating contextual factors in technology-based solutions designed to support health and fitness routines for older adults: A systematic review. In *Human-Computer Interaction. Design and Research*, page 161–192. Springer, Cham, 2025. ISBN 978-3-031-80829-6. doi:10.1007/978-3-031-80829-6_8. URL https://link.springer.com/chapter/10.1007/978-3-031-80829-6_8.

[155] F. Regazzoni, P. Palmieri, F. Smailbegovic, R. Cammarota, and I. Polian. Protecting artificial intelligence ips: a survey of watermarking and fingerprinting for machine learning. *CAAI Transactions on Intelligence Technology*, 6(2):180–191, 2021. ISSN 2468-2322. doi:10.1049/cit2.12029.

[156] A. Reisner. These 183,000 books are fueling the biggest fight in publishing and tech. *The Atlantic*, Sept. 2023. URL https://www.theatlantic.com/technology/archive/2023/09/books3-database-generative-ai-training-copyright-infringement/675363/.

[157] J. M. Rigby and C. Preist. Towards user-centred climate services: the role of human-computer interaction. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, page 1–14, New York, NY, USA, Apr. 2023. Association for Computing Machinery. ISBN 978-1-4503-9421-5. doi:10.1145/3544548.3580663. URL https://dl.acm.org/doi/10.1145/3544548.3580663.

[158] D. K. Rosner. *Critical Fabulations: Reworking the Methods and Margins of Design*. The MIT Press, June 2018.

[159] D. Rotman, J. Preece, J. Hammock, K. Procita, D. Hansen, C. Parr, D. Lewis, and D. Jacobs. Dynamic changes in motivation in collaborative citizen-science projects. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, CSCW '12, page 217–226, New York, NY, USA, Feb 2012. Association for Computing Machinery. ISBN 978-1-4503-1086-4. doi:10.1145/2145204.2145238. URL https://doi.org/10.1145/2145204.2145238.

[160] D. Russo. Navigating the complexity of generative ai adoption in software engineering. *ACM Trans. Softw. Eng. Methodol.*, 33(5):135:1–135:50, June 2024. ISSN 1049-331X. doi:10.1145/3652154.

[161] M. Sako. How generative ai fits into knowledge work. *Commun. ACM*, 67(4):20–22, Mar. 2024. ISSN 0001-0782. doi:10.1145/3638567.

[162] J. Saldañna. *The Coding Manual for Qualitative Researchers*. SAGE, Los Angeles, 2nd edition, 2013. ISBN 978-1-4462-4736-5 978-1-4462-4737-2.

[163] S. A. Scherr, A. Meier, and S. Cihan. Alexa, tell me more – about new best friends, the advantage of hands-free operation and life-long learning. page 10.18420/muc2020. Gesellschaft für Informatik e.V., 2020. URL https://dl.gi.de/handle/20.500.12116/33523.

[164] T. Schick, J. Dwivedi-Yu, R. Dessì, R. Raileanu, M. Lomeli, E. Hambro, L. Zettlemoyer, N. Cancedda, and T. Scialom. Toolformer: Language models can teach themselves to use tools. In *Advances in Neural Information Processing Systems*, NeurIPS '23, pages 68539–68551, 2023. ISBN 978-1-4503-9421-5.

[165] D. Schien, P. Shabajee, J. Wickenden, W. Picket, G. Roberts, and C. Preist. Demo: The dimpact tool for environmental assessment of digital services. In *ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies (COMPASS)*, COMPASS '22, page 701–703, New York, NY, USA, Jun 2022. Association for Computing Machinery. ISBN 978-1-4503-9347-8. doi:10.1145/3530190.3542932. URL https://doi.org/10.1145/3530190.3542932.

[166] A. Schlesinger, K. P. O'Hara, and A. S. Taylor. Let's talk about race: Identity, chatbots, and ai. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, page 1–14, New York, NY, USA, Apr. 2018. Association for

Computing Machinery. ISBN 978-1-4503-5620-6. doi:10.1145/3173574.3173889. URL https://dl.acm.org/doi/10.1145/3173574.3173889.

[167] J. Q. Schmidt. Climate resiliency through data: Managing stormwater sewers in detroit. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI EA '22, page 1–5, New York, NY, USA, Apr 2022. Association for Computing Machinery. ISBN 978-1-4503-9156-6. doi:10.1145/3491101.3516804. URL https://doi.org/10.1145/3491101.3516804.

[168] Y. S. Sefidgar, C. L. Castillo, S. Chopra, L. Jiang, T. Jones, A. Mittal, H. Ryu, J. Schroeder, A. Cole, N. Murinova, et al. Migrainetracker: Examining patient experiences with goal-directed self-tracking for a chronic health condition. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pages 1–19, 2024. doi:10.1145/3613904.3642075.

[169] A. Sellen and E. Horvitz. The rise of the ai co-pilot: Lessons for design from aviation and beyond. *Commun. ACM*, 67(7):18–23, July 2024. ISSN 0001-0782. doi:10.1145/3637865.

[170] K. Sengupta, S. Sarcar, A. Pradhan, R. McNaney, S. Sayago, D. Chattopadhyay, and A. Joshi. Challenges and opportunities of leveraging intelligent conversational assistant to improve the well-being of older adults. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI EA '20, page 1–4, New York, NY, USA, Apr 2020. Association for Computing Machinery. ISBN 978-1-4503-6819-3. doi:10.1145/3334480.3381057. URL https://doi.org/10.1145/3334480.3381057.

[171] E. Sezgin, L. K. Militello, Y. Huang, and S. Lin. A scoping review of patient-facing, behavioral health interventions with voice assistant technology targeting self-management and healthy lifestyle behaviors. *Translational Behavioral Medicine*, 10(3):606–628, Aug. 2020. ISSN 1869-6716, 1613-9860. doi:10.1093/tbm/ibz141.

[172] S. Shan, W. Ding, J. Passananti, S. Wu, H. Zheng, and B. Y. Zhao. Nightshade: Prompt-specific poisoning attacks on text-to-image generative models. page 807–825. IEEE Computer Society, May 2024. ISBN 9798350331301. doi:10.1109/SP54263.2024.00207. URL https://www.computer.org/csdl/proceedings-article/sp/2024/313000a212/1WPcYDRkVX2.

[173] E. Shandilya and M. Fan. Understanding older adults' perceptions and challenges in using ai-enabled everyday technologies. In *Proceedings of the Tenth International Symposium of Chinese CHI*, Chinese CHI '22, page 105–116, New York, NY, USA, Feb. 2024. Association for Computing Machinery. ISBN 978-1-4503-9869-5. doi:10.1145/3565698.3565774. URL https://dl.acm.org/doi/10.1145/3565698.3565774.

[174] J. Shi, R. Jain, H. Doh, R. Suzuki, and K. Ramani. An hci-centric survey and taxonomy of human-generative-ai interactions. (arXiv:2310.07127), Jan. 2024. doi:10.48550/arXiv.2310.07127. URL `http://arxiv.org/abs/2310.07127`.

[175] B. Shneiderman. *Human-Centered AI.* Oxford University Press, Oxford, Feb. 2022. ISBN 978-0-19-284529-0.

[176] I. Siles, A. Segura-Castillo, R. Solís, and M. Sancho. Folk theories of algorithmic recommendations on spotify: Enacting data assemblages in the global south. *Big Data Society*, 7(1):2053951720923377, Jan 2020. ISSN 2053-9517. doi:10.1177/2053951720923377.

[177] N. Silver. *The Signal and the Noise: Why So Many Predictions Fail–but Some Don't.* Penguin Books, New York, NY, illustrated edition edition, Feb. 2015. ISBN 978-0-14-312508-2.

[178] A. Singh, A. Klapper, J. Jia, A. Fidalgo, A. Tajadura-Jiménez, N. Kanakam, N. Bianchi-Berthouze, and A. Williams. Motivating people with chronic pain to do physical activity: Opportunities for technology design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, page 2803–2812, New York, NY, USA, 2014. Association for Computing Machinery. ISBN 9781450324731. doi:10.1145/2556288.2557268. URL `https://doi.org/10.1145/2556288.2557268`.

[179] G. R. Smith, C. Bello, L. Bialic-Murphy, E. Clark, C. S. Delavaux, C. Fournier de Lauriere, J. van den Hoogen, T. Lauber, H. Ma, D. S. Maynard, M. Mirman, L. Mo, D. Rebindaine, J. E. Reek, L. K. Werden, Z. Wu, G. Yang, Q. Zhao, C. M. Zohner, and T. W. Crowther. Ten simple rules for using large language models in science, version 1.0. *PLOS Computational Biology*, 20(1), 2024. doi:10.1371/journal.pcbi.1011767.

[180] J. Snyder. Vernacular visualization practices in a citizen science project. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '17, page 2097–2111, New York, NY, USA, Feb 2017. Association for Computing Machinery. ISBN 978-1-4503-4335-0. doi:10.1145/2998181.2998239. URL `https://doi.org/10.1145/2998181.2998239`.

[181] J. So, S. Estrada, M. Jörke, E. Bianchi, M. Wang, N. Haghighi, K. L. Fessele, J. A. Landay, and A. Cuadra. "they make us old before we're old": Designing ethical health technology with and for older adults. *Proc. ACM Hum.-Comput. Interact.*, 8(CSCW2): 478:1–478:30, Nov. 2024. doi:10.1145/3687017.

[182] S. L. Song, B. Kruft, M. Zhang, C. Li, S. Chen, C. Zhang, M. Tanaka, X. Wu, J. Rasley, A. A. Awan, C. Holmes, M. Cai, A. Ghanem, Z. Zhou, Y. He, P. Luferenko, D. Kumar, J. Weyn, R. Zhang, S. Klocek, V. Vragov, M. AlQuraishi, G. Ahdritz, C. Floristean, C. Negri, R. Kotamarthi, V. Vishwanath, A. Ramanathan, S. Foreman, K. Hippe, T. Arcomano, R. Maulik, M. Zvyagin, A. Brace, B. Zhang, C. O. Bohorquez, A. Clyde, B. Kale, D. Perez-Rivera, H. Ma, C. M. Mann, M. Irvin, J. G. Pauloski,

L. Ward, V. Hayot, M. Emani, Z. Xie, D. Lin, M. Shukla, I. Foster, J. J. Davis, M. E. Papka, T. Brettin, P. Balaprakash, G. Tourassi, J. Gounley, H. Hanson, T. E. Potok, M. L. Pasini, K. Evans, D. Lu, D. Lunga, J. Yin, S. Dash, F. Wang, M. Shankar, I. Lyngaas, X. Wang, G. Cong, P. Zhang, M. Fan, S. Liu, A. Hoisie, S. Yoo, Y. Ren, W. Tang, K. Felker, A. Svyatkovskiy, H. Liu, A. Aji, A. Dalton, M. Schulte, K. Schulz, Y. Deng, W. Nie, J. Romero, C. Dallago, A. Vahdat, C. Xiao, T. Gibbs, A. Anandkumar, and R. Stevens. Deepspeed4science initiative: Enabling large-scale scientific discovery through sophisticated ai system technologies. (arXiv:2310.04610), Oct. 2023. doi:10.48550/arXiv.2310.04610. URL http://arxiv.org/abs/2310.04610.

[183] S. L. Star. This is not a boundary object: Reflections on the origin of a concept. *Science, technology, & human values*, 35(5):601–617, 2010.

[184] B. Stigall, J. Waycott, S. Baker, and K. Caine. Older adults' perception and use of voice user interfaces: A preliminary review of the computing literature. In *Proceedings of the 31st Australian Conference on Human-Computer-Interaction*, OzCHI '19, page 423–427, New York, NY, USA, Jan. 2020. Association for Computing Machinery. ISBN 978-1-4503-7696-9. doi:10.1145/3369457.3369506. URL https://dl.acm.org/doi/10.1145/3369457.3369506.

[185] M. Sturken. Desiring the weather: El nino, the media, and california identity. *Public Culture*, 13(2):161–189, 2001. ISSN 1527-8018.

[186] M. Sturken. Seeing the temperature in weather media. 2014.

[187] L. A. Suchman. Plans and situated actions - the problem of human-machine communication. In *Learning in doing: social,cognitive,and computational perspectives*, 1987. doi:10.2307/2073874.

[188] M. M. Suh, E. Youngblom, M. Terry, and C. J. Cai. Ai as social glue: Uncovering the roles of deep generative ai during social music composition. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, page 1–11, New York, NY, USA, May 2021. Association for Computing Machinery. ISBN 978-1-4503-8096-6. doi:10.1145/3411764.3445219. URL https://dl.acm.org/doi/10.1145/3411764.3445219.

[189] S. Suh, B. Min, S. Palani, and H. Xia. Sensecape: Enabling multilevel exploration and sensemaking with large language models. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, UIST '23, page 1–18, New York, NY, USA, Oct. 2023. Association for Computing Machinery. ISBN 9798400701320. doi:10.1145/3586183.3606756. URL https://dl.acm.org/doi/10.1145/3586183.3606756.

[190] Y. Sun, E. Jang, F. Ma, and T. Wang. Generative ai in the wild: Prospects, challenges, and strategies. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI '24, page 1–16, New York, NY, USA, May 2024. Association

for Computing Machinery. ISBN 9798400703300. doi:10.1145/3613904.3642160. URL `https://dl.acm.org/doi/10.1145/3613904.3642160`.

[191] H. Suresh, R. Movva, A. L. Dogan, R. Bhargava, I. Cruxen, A. M. Cuba, G. Taurino, W. So, and C. D'Ignazio. Towards intersectional feminist and participatory ml: A case study in supporting feminicide counterdata collection. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '22, page 667–678, New York, NY, USA, June 2022. Association for Computing Machinery. ISBN 978-1-4503-9352-2. doi:10.1145/3531146.3533132. URL `https://dl.acm.org/doi/10.1145/3531146.3533132`.

[192] M. Sänger, N. De Mecquenem, K. E. Lewińska, V. Bountris, F. Lehmann, U. Leser, and T. Kosch. A qualitative assessment of using chatgpt as large language model for scientific workflow development. *GigaScience*, 13:1–19, 2024. doi:10.1093/gigascience/giae030.

[193] X. Tang, Y. Sun, B. Zhang, Z. Liu, R. LC, Z. Lu, and X. Tong. "i never imagined grandma could do so well with technology": Evolving roles of younger family members in older adults' technology learning and use. *Proc. ACM Hum.-Comput. Interact.*, 6 (CSCW2):478:1–478:29, Nov. 2022. doi:10.1145/3555579.

[194] R. Taylor, M. Kardas, G. Cucurull, T. Scialom, A. Hartshorn, E. Saravia, A. Poulton, V. Kerkez, and R. Stojnic. Galactica: A large language model for science. (arXiv:2211.09085), Nov. 2022. doi:10.48550/arXiv.2211.09085. URL `http://arxiv.org/abs/2211.09085`.

[195] M. Trajkova and A. Martin-Hammond. *"Alexa is a Toy:" Exploring Older Adults' Reasons for Using, Limiting, and Abandoning Echo*, page 1–13. Association for Computing Machinery, New York, NY, USA, Apr 2020. ISBN 978-1-4503-6708-0. URL `https://doi.org/10.1145/3313831.3376760`.

[196] P. Upadhyay, S. Heung, S. Azenkot, and R. N. Brewer. Studying exploration & long-term use of voice assistants by older adults. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9781450394215. doi:10.1145/3544548.3580925. URL `https://doi.org/10.1145/3544548.3580925`.

[197] T. Usher, B. Buta, J. Thorpe, Roland J, J. Huang, L. J. Samuel, J. D. Kasper, and K. Bandeen-Roche. Dissecting the racial/ethnic disparity in frailty in a nationally representative cohort study with respect to health, income, and measurement. *The Journals of Gerontology: Series A*, 76(1):69–76, Jan 2021. ISSN 1079-5006. doi:10.1093/gerona/glaa061.

[198] S. Uusitalo, A. Salovaara, T. Jokela, and M. Salmimaa. "clay to play with": Generative ai tools in ux and industrial design practice. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference*, DIS '24, page 1566–1578, New York,

NY, USA, July 2024. Association for Computing Machinery. ISBN 9798400705830. doi:10.1145/3643834.3661624. URL https://dl.acm.org/doi/10.1145/3643834.3661624.

[199] D. Vargemidis, K. Gerling, V. V. Abeele, L. Geurts, and K. Spiel. Irrelevant Gadgets or a Source of Worry: Exploring Wearable Activity Trackers with Older Adults. *ACM Trans. Access. Comput.*, 14(3):16:1–16:28, Aug. 2021. ISSN 1936-7228. doi:10.1145/3473463.

[200] K. B. Wagman. *Sex, Power, and Technology: A Relational Engineering Ethos as Feminist Utopia*. Thesis, Massachusetts Institute of Technology, June 2021. URL https://dspace.mit.edu/handle/1721.1/139290.

[201] K. B. Wagman and L. Parks. Beyond the command: Feminist STS research and critical issues for the design of social machines. *Proc. ACM Hum.-Comput. Interact.*, 5(CSCW1), Apr. 2021. doi:10.1145/3449175. URL https://doi.org/10.1145/3449175.

[202] K. B. Wagman, M. T. Dearing, and M. Chetty. Generative AI uses and risks for knowledge workers in a science organization. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, CHI '25, New York, NY, USA, 2025. Association for Computing Machinery. ISBN 9798400713941. doi:10.1145/3706598.3713827. URL https://doi.org/10.1145/3706598.3713827.

[203] A. M. Walker, Y. Yao, C. Geeng, R. Hoyle, and P. Wisniewski. Moving beyond 'one size fits all:' research considerations for working with vulnerable populations. *Interactions*, 26(6):34–39, Oct. 2019. ISSN 1072-5520, 1558-3449. doi:10.1145/3358904.

[204] J. Wang, C. Huang, S. Yan, W. Xie, and D. He. When young scholars cooperate with llms in academic tasks: The influence of individual differences and task complexities. *International Journal of Human–Computer Interaction*, page 1–16, 2024. doi:10.1080/10447318.2024.2352919.

[205] R. Wang, R. Cheng, D. Ford, and T. Zimmermann. Investigating and designing for trust in ai-powered code generation tools. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '24, page 1475–1493, New York, NY, USA, June 2024. Association for Computing Machinery. ISBN 9798400704505. doi:10.1145/3630106.3658984. URL https://dl.acm.org/doi/10.1145/3630106.3658984.

[206] Y. Wang, A. Fadhil, and H. Reiterer. Supporting action planning for sedentary behavior change by visualizing personal mobility patterns on smartphone. In *Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare*, PervasiveHealth '18, page 396–401, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450364508. doi:10.1145/3240925.3240962. URL https://doi.org/10.1145/3240925.3240962.

[207] Y. Wang, L. M. König, H. Reiterer, et al. A smartphone app to support sedentary behavior change by visualizing personal mobility patterns and action planning (sedvis): Development and pilot study. *JMIR formative research*, 5(1):e15369, 2021. doi:10.2196/15369.

[208] Y. Wang, M. Li, Y.-H. Kim, B. Lee, M. Danilovich, A. Lazar, D. E. Conroy, H. Kacorri, and E. K. Choe. Redefining Activity Tracking Through Older Adults' Reflections on Meaningful Activities. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24, pages 1–15. Association for Computing Machinery, May 2024. ISBN 979-8-4007-0330-0. doi:10.1145/3613904.3642170.

[209] M. J. Western, M. E. Armstrong, I. Islam, K. Morgan, U. F. Jones, and M. J. Kelson. The effectiveness of digital interventions for increasing physical activity in individuals of low socioeconomic status: A systematic review and meta-analysis. *International Journal of Behavioral Nutrition and Physical Activity*, 18:1–21, 2021. doi:10.1186/s12966-021-01218-4.

[210] N. Wong, S. Jeong, M. Reddy, C. A. Stamatis, E. G. Lattie, and M. Jacobs. Voice assistants for mental health services: Designing dialogues with homebound older adults. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference*, DIS '24, page 844–858, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400705830. doi:10.1145/3643834.3661536. URL `https://doi.org/10.1145/3643834.3661536`.

[211] A. Woodruff, R. Shelby, P. G. Kelley, S. Rousso-Schindler, J. Smith-Loud, and L. Wilcox. How knowledge workers think generative ai will (not) transform their industries. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI '24, page 1–26, New York, NY, USA, May 2024. Association for Computing Machinery. ISBN 9798400703300. doi:10.1145/3613904.3642700. URL `https://doi.org/10.1145/3613904.3642700`.

[212] K. Xu, X. Yan, and M. W. Newman. Understanding people's experience for physical activity planning and exploring the impact of historical records on plan creation and execution. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450391573. doi:10.1145/3491102.3501997. URL `https://doi.org/10.1145/3491102.3501997`.

[213] K. Xu, X. E. Yan, M. Ryu, M. W. Newman, and R. I. Arriaga. Understanding the effect of reflective iteration on individuals' physical activity planning. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400703300. doi:10.1145/3613904.3641937. URL `https://doi.org/10.1145/3613904.3641937`.

[214] W. Xu. Toward human-centered ai: a perspective from human-computer interaction. *interactions*, 26(4):42–46, June 2019. ISSN 1072-5520. doi:10.1145/3328485.

[215] Q. Yang, A. Steinfeld, and J. Zimmerman. Unremarkable ai: Fitting intelligent decision support into critical, clinical decision-making processes. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, page 1–11, New York, NY, USA, May 2019. Association for Computing Machinery. ISBN 978-1-4503-5970-2. doi:10.1145/3290605.3300468. URL https://doi.org/10.1145/3290605.3300468.

[216] N. Zargham, D. Alexandrovsky, J. Erich, N. Wenig, and R. Malaka. "i want it that way": Exploring users' customization and personalization preferences for home assistants. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI EA '22, page 1–8, New York, NY, USA, Apr. 2022. Association for Computing Machinery. ISBN 978-1-4503-9156-6. doi:10.1145/3491101.3519843. URL https://dl.acm.org/doi/10.1145/3491101.3519843.

[217] R. Ziman and G. Walsh. Factors affecting seniors' perceptions of voice-enabled user interfaces. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI EA '18, page 1–6, New York, NY, USA, Apr 2018. Association for Computing Machinery. ISBN 978-1-4503-5621-3. doi:10.1145/3170427.3188575. URL https://doi.org/10.1145/3170427.3188575.

[218] T. Zubatiy, K. L. Vickers, N. Mathur, and E. D. Mynatt. Empowering dyads of older adults with mild cognitive impairment and their care partners using conversational agents. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, page 1–15, New York, NY, USA, May 2021. Association for Computing Machinery. ISBN 978-1-4503-8096-6. doi:10.1145/3411764.3445124. URL https://doi.org/10.1145/3411764.3445124.

[219] T. Zubatiy, N. Mathur, L. Heck, K. L. Vickers, A. Rozga, and E. D. Mynatt. "i don't know how to help with that" – learning from limitations of modern conversational agent systems in caregiving networks. *Proc. ACM Hum.-Comput. Interact.*, 7(CSCW2), Oct. 2023. doi:10.1145/3610170. URL https://doi.org/10.1145/3610170.