Manuscript ID 2022-11-0089-OA10-TTS

# Ethical Issues in Near-Future Socially Supportive Smart Assistants for Older Adults

Alex John London, Yosef S. Razin, Jason Borenstein, Motahhare Eslami, Russell Perkins, and Paul Robinette

*Abstract*— **This paper considers novel ethical issues pertaining to near-future artificial intelligence (AI) systems that seek to support, maintain, or enhance the capabilities of older adults as they age and experience cognitive decline. In particular, we focus on smart assistants (SAs) that would seek to provide proactive assistance and mediate social interactions between users and other members of their social or support networks. Such systems would potentially have significant utility for users and their caregivers if they could reduce the cognitive load for tasks that help older adults maintain their autonomy and independence. However, proactively supporting even simple tasks, such as providing the user with a summary of a meeting or a conversation, would require a future SA to engage with ethical aspects of human interactions which computational systems currently have difficulty identifying, tracking, and navigating. If SAs fail to perceive ethically relevant aspects of social interactions, the resulting deficit in moral discernment would threaten important aspects of user autonomy and well-being. After describing the dynamic that generates these ethical challenges, we note how simple strategies for prompting user oversight of such systems might also undermine their utility. We conclude by considering how near-future SAs could exacerbate current worries about privacy, commodification of users, trust calibration and injustice.**

*Index Terms*—**Ethics, moral discernment, dependence, vulnerability, assistive technology, older adults, mild cognitive impairment, smart assistants**

## I. INTRODUCTION

A longstanding ambition of Artificial Intelligence (AI) research has been to develop assistive technologies that can support, maintain, and enhance the capabilities of people as they age or face various sources of physical, affective, or cognitive decline. The proliferation of smart assistants (SAs), such as Amazon's Alexa, Microsoft's Cortana, and Google Assistant, that can integrate with smart home technology and Internet of Things (IOT) devices, suggests that future iterations of SAs have the potential to play an important role within a larger AI support system. The ability of users, and their caregivers, to interact with these systems through voice commands rather than keystrokes may make them particularly attractive candidates for use with older adults or others experiencing various types of cognitive or physical decline [1]. Such devices might also serve as the central hub or "brain" for future technologies that integrate services beyond the confines of the home, helping to mediate social interactions between users and elements of a wider social support network. As users travel to appointments with family, friends, or other caregivers, for example, their SA might travel with them on a mobile phone or smart wearable device. Future SAs may thus be able to incorporate data from a wide range of sensors across a wide range of domains, both within and outside of the home. Long-term use of SAs creates the possibility for longitudinal learning in which a nuanced model of the user is constructed and refined, including their relationships within social and support networks. This capability, among others, creates the potential for SAs that go beyond detecting acute events, such as a fall or heart attack, to identifying gradual declines in the user's cognitive and motor abilities [2]. If SAs could detect declines in user abilities and dynamically adapt to support user needs, this would make significant strides in achieving one of the AI community's oldest dreams— enabling individuals to function in ways that preserve their autonomy while simultaneously promoting and protecting their personal well-being.

Yet, the ambition of developing highly interactive SAs that anticipate user needs and mediate interactions in a larger social network raises multifaceted, complex, and novel ethical issues. We argue that these issues are rooted in three interrelated dynamics.

First, when older adults rely on AI systems to maintain their autonomy and support their wellbeing, such adults become vulnerable in unique ways. Systems that fail to perform required functions at the right time, or in the right way, leave older adults vulnerable to compromises in autonomy or welfare that might have been avoided had they chosen alternative means of assistance.

Second, current SAs are reactive in the sense that they rely on users to perform key cognitive tasks, such as identifying a use case for the system, scaffolding how the system can achieve user goals, and then initiating tasks required to effectuate this

This work was supported by the U.S. National Science Foundation Grant 2112633. *Corresponding author: Alex John London.* London is the lead author on the manuscript; authors are ordered in terms of the significance of their intellectual contribution.

Alex John London is with the Department of Philosophy, Carnegie Mellon University, Pittsburgh, PA 15213 USA (e-mail: ajlondon@andrew.cmu.edu).

Yosef S. Razin is with the School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA (e-mail: yrazin@gatech.edu).

Jason Borenstein is with the School of Public Policy and Office of Graduate Education, Georgia Institute of Technology, Atlanta, GA (e-mail: borenstein@gatech.edu).

Motahhare Eslami is with the Human-Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, PA 15213, USA (e-mail: meslami@andrew.cmu.edu).

Russell Perkins is with the Department of Electrical and Computer Engineering, University of Massachusetts Lowell (e-mail: Russell_Perkins@student.uml.edu).

Paul Robinette is with the Department of Electrical and Computer Engineering, University of Massachusetts Lowell, (e-mail: Paul_Robinette@uml.edu).

plan. To be proactive, future SAs will need to take on some of these cognitive tasks. But aiding with what might appear to be a relatively simple cognitive task, such as providing the user with a summary of a meeting or a conversation, requires SAs to engage with ethical aspects of human interactions which computational systems currently have difficulty identifying, tracking, and navigating. Failure to perceive ethically relevant aspects of social interactions constitutes a deficit in moral discernment that threatens aspects of user autonomy and well-being. Ambiguities within language and complexities in how language is used to communicate beyond literal assertion is one among many challenges that designers will have to overcome [3].

Third, current SAs function in dyadic relationships with users or mediate relationships between users and smart devices. To mediate social relationships with parties that provide social services, other members of their care team, or family and friends, future SAs will have to be able to navigate more complex and ethically laden aspects of the social world. Delegating tasks in the social world to SAs requires that such systems can ascertain the structure of moral relationships and act in ways that respect a network of expectations, rights, duties, and permissions. Failures in this space can also have profound consequences for user autonomy and well-being.

Efforts to manage these vulnerabilities raise additional ethical issues. In particular, whether a future SA can function in ways that provide a net benefit to the user hinges on its ability to perform tasks that advance the user's projects and plans without requiring tedious or complex oversight or extensive auditing of its performance. The ambition of providing proactive assistance or mediating social relationships increases the challenge of demarcating which tasks an SA can perform and communicating the conditions under which it can perform those tasks reliably. This difficulty is exacerbated by the prospect that the users most in need of such assistance are those at risk of, or already experiencing, cognitive decline.

Similarly, to respect user autonomy and to ensure that an SA is truly an assistant—that it is providing support for the goals and ends of the user rather than manipulating the user to pursue someone else's goals and ends —care must be taken to avoid conditioning users to change their behavior around the capabilities of the system. It will also require a robust and highly granular model of user goals, preferences, and capabilities. Constructing such model would not only entail capturing large volumes of sensitive, private information, it would likely involve generating new information about the user, such as projecting the rate at which further cognitive decline may occur to better anticipate when and which types of additional assistance will be necessary. However, the prospect that future SAs would combine vast troves of longitudinal, multi-modal information in ways that might create new sensitive information about the user raises profound privacy and confidentiality issues. If systems are to be granted access to the most sensitive and private aspects of a person's life, then there is a strong moral case that they should be designed and function as fiduciaries of the user's interests. This has important implications for the business model that might support the development of such assistants.

Finally, the goal of creating assistive systems that can tailor their activities to the capabilities and values of the individual user raises difficult ethics issues related to fairness and equity. Such systems will need to adapt not only to variation in speech patterns across categories such as gender and geographic origin, but they will have to navigate speech patterns that may arise for users with medical conditions that impair their ability to communicate.

In the following sections, we review some of the ethics literature relevant to near-future smart assistants, with a particular focus on older adults experiencing mild cognitive impairment (MCI). We then provide examples of tasks that near-future SAs might undertake to provide proactive assistance in a social space, highlighting some of the unique ethical challenges that arise from these ambitions. To make an already broad topic more manageable, we focus specifically on cases in which the person in need of social support is also the party who purchases and operates the SA. Circumstances in which SAs are employed by third parties, such as children or caregivers, to monitor or assist a loved one will be the subject of future work.

## II. The Ethical Landscape Around Smart Assistants

Near-future SAs of the sort that we consider in this paper would likely raise the full range of ethical issues that have already been discussed in the context of current SAs, assistive technologies [4], smart homes [5], IoT [6], and ubiquitous monitoring [7], [8], network security and surveillance [6], [7], [9], and multi-modal sensing including inputs from wearable and other devices [4][10][11]. There is also a growing literature that surveys the broad range of ethical issues that have emerged in relation to AI, including reports by IEEE [12] and various governmental and non-governmental entities [13], [14], [15], [16], [17] . The issues include, but are not limited to, safety and security, explainability, fairness and non-discrimination, human control of technology, and the professional responsibilities of technology designers. An important and growing thread within the literature is the alignment of AI with human values [16], [17]. Many of these issues will take on new importance in the context of SAs that seek to be proactive and to mediate social interactions. As a result, we discuss below how these novel features of SAs intersect with issues of privacy, trust, agency, and control along with highlighting some of the current literature on such topics.

### *A. Human Values and Identity*

Prior work identifies significant ethical challenges that emerge from a user's relationship with assistive technology over time. Emotional attachment can facilitate technological acceptance and maintain user trust [18] while creating the potential for user over-trust, manipulation, and emotional dependence [19], [20]. Emotional injury can result from a technology's inability to maintain genuine affective relationships and inability to provide true recognition of their

users. Autonomous systems that provide a false sense of recognition or perception can undermine user identity and integrity and exacerbate social isolation in vulnerable populations [19], [20]. Prior work also notes the potential for mechanization and standardization to undermine personhood and identity and to compound inequality and injustice [19]. Our work builds on these concerns by highlighting the complexities associated with ensuring that future SAs advance the projects and plans of users without promoting overtrust or nudging users to adjust their goals and plans to conform to the abilities of assistive systems.

### *B. Proactivity, Agency, and Control*

A major gap in the literature that provides an impetus for our work is the ethics of proactivity, which has received little attention in the AI and computing community. Proactivity, perhaps more than any other feature of near-future SAs, is premised on properly anticipating and projecting a user's future goals, preferences, priorities, and plans. Proactive systems in private spaces epitomize the need for personalization and open-ended design, where systems afford users multiple modes of interaction [5]. At the same time as proactive system designers create more space for the user and their goals, they also attempt to reduce the need for direct control and monitoring by the user. This introduces a 'responsibility gap' between the technology and its developers, where the user, either as teacher, subject, or owner of the system, is put in a position of increased responsibility for deciding to even use the system [21]. Just as a 'gap' develops between designers and users, another forms between the user and the system itself. Known in the literature as the 'responsibility-authority double bind' [22], the system's decision-making responsibility increases as the user cedes control, putting further weight on the user to monitor, understand, and trust such systems.

Furthermore, as proactivity reduces the need for direct control, it can result in users losing their sense of agency and becoming more vulnerable [23], [5]. The loss of control can result in an uncanny sense of alienation, especially if compounded by suspicions of surveillance [23]. On the other hand, relatively unintrusive or invisible systems can cause one to forget they are present [23], which may imperil privacy and consent. Finally, the outsourcing and delegation of tasks to a technological system can lead to the degradation of a user's attention, engagement, and skills [24].

### *C. Trust*

The vulnerabilities associated with reliance on a socially and proactively supportive SA make trust particularly important issue for future SAs. We build on prior work that understands trust as a set of attitudes, beliefs, and intentions about voluntary delegation of one's goals [25]. Trust between humans and automated systems has been explored in a variety of ways over the last few decades [26], [27], [28]. Consideration of delegation is premised on the user's expectation of the system's cooperativeness, capability, and constraints [29], [30], [31]– not only that it can do the necessary tasks but whether it will behave within the ethical norms or laws that the user anticipates as well [32], [33]. These expectations may arise from general attitudes toward technology, the user's own trusting stance toward the world at large, or factors that are more specific to the system, such as its reputation, understandability, and the quality of its recommendations [34].

Expectations like these can lead to overtrust, where users trust the technology more than its goals and capabilities warrant [35], [36]. Initial expectations of machines may assume they will perform perfectly until proven otherwise [37] and robots and AI are often to assumed to be fairer [38], more impartial, and legitimate [39].

Beyond the previously mentioned concerns, trust in SAs for those with MCI is particularly challenging. Depending on how MCI manifests itself, one's ability to discern and set goals, may be impaired [40]. These issues and their implications for design have yet to be explored for SAs, much less proactive ones that are highly personalized and have the ability to cause emotional injury if misused.

### *D. Privacy*

The amount and types of data that assistive technology should collect, share, and store have been frequent topics of discussion [9]. Privacy concerns extend beyond physical behaviors and formal policies and include identifying system vulnerabilities, threats to safety and security [41], appropriate transparency and control, and recognizing differences in data sensitivity [8]. Currently available systems do not conform to transparency standards around privacy [8] and the techniques used by these systems, such as nudging [42] or hypernudging (personalized behavioral influences), puts data and assistance into such a tightly coupled loop between the user and a wide variety of support services, that understandability of data usage deteriorates significantly [43]. These issues are only compounded for future SAs that seek to gather data both inside and outside the home, in sensitive personal and social contexts, across multiple modalities.

## III. Near Future Socially Supportive Smart Assistants

Current SAs can carry out a limited range of tasks to support older users. These include facilitating information recall through voice inputs, storage, and retrieval of information [44]. They can also be used to control smart appliances and to initiate communication through phone calls or text messages. These systems are limited, however, by two factors.

First, current systems rely on users to carry out complex planning and perceptual tasks. If a user deems an event to be important, they can use the SA to set a reminder and to speak or display an alert. Frequently, however, the timing and content of the alert must be determined by the user. If an event requires a sequence of steps, the user must understand this sequence, formulate a plan for carrying it out, and then determine how to incorporate the SA into this plan. For example, they might tell their SA to set a reminder at some point before each step and then include information about subsequent steps.

Second, older adults often experience declines in memory,

sensory, and cognitive abilities, such as hearing, visuoperceptual judgment, speech comprehension and verbal fluency/retrieval [45], [46] in addition to changes in attention and executive function, including decision making and judgment [45], [47]. The cumulative effect of even mild impairments in these abilities can create significant impediments to function independently and to carry out tasks that play an important role in a person's sense of self or for finding an outlet for well-being and fulfillment [19].

To make the fullest use of current SAs, older adults who experience cognitive decline, such as those with MCI, must engage in the second-order task of anticipating their future needs, formulating a plan around the SA's capabilities, and then engaging with the SA to implement that plan in practice. However, the very cognitive faculties whose decline the SA is intended to overcome are simultaneously required to make full use of such systems. Users who are prone to forgetfulness or who have difficulty carrying out a complex series of tasks without assistance might want to use an SA for support, but the process of configuring the SA to advance their interests requires users to draw on these same cognitive capacities whose shortcomings necessitated the use of the system's supportive features in the first place. Family or caregivers can reduce this burden by helping with this process, but not all older adults have this option, and even those who do may see such reliance on caregivers as a burden, an intrusion, or a compromise to their autonomy or independence.

A potential pathway to increase the utility of SAs would be to provide those systems with the capabilities necessary to be more proactive in supporting older adults. An SA that has a granular profile of the user's needs and values might anticipate events with which the user might require support. The ambition would be to reduce the cognitive load on users by anticipating user needs and proactively acting to support their goals and purposes without the need for users to initiate or to scaffold this process.

To make these ideas concrete, we consider two tasks for which a future SA might provide proactive support. First, advances in natural language processing (NLP) have generated interest in systems that support information collection and retrieval, not by creating a transcript of a complete conversation or requiring the user to input a summary, but by autonomously creating a condensed summary of key information [48], [49], [50]. A system that could extract and record relevant information from routine social interactions (e.g., conversations with friends, doctor's visits, interactions with service providers or caregivers), without the user having to input that information directly, might enable older adults to manage social commitments, meet social expectations, or take advantage of social opportunity. Such a system would provide proactive assistance by taking over responsibility for an important cognitive task—identifying and recording salient information for later retrieval by the user.

Second, systems could support older adults by taking on some of the cognitive load associated with planning and scheduling. This would involve using information known to be important to the user to scaffold reminders and plan the series of tasks and prompts necessary for users to effectuate relevant goals and commitments.

Using language from Clark and Chalmers [51], proactive SAs would extend the minds of users by augmenting their capacity not just to store and retrieve information but to collect it and to organize it into plans that support their goals and ends. Their utility for older adults facing cognitive and physical decline lies in the prospect that they might augment the very cognitive, perceptual, and agentic capacities that are impaired by age and cognitive decline.

The unique ethical challenge facing such systems is that each of these cognitive, perceptual and agentic tasks is morally laden—they require an awareness of what features of an interaction are important, their degree of relative importance, their connection to other moral values, commitments, obligations or responsibilities, and a sense of how to interact with the world in a way that is responsive to these moral norms or requirements. We now illustrate these claims in more concrete detail.

## IV. Context and Moral Discernment

Information gathering, synthesis, and planning require the ability to identify and respond to morally relevant features of the world. We say that a lack of moral discernment occurs when an agent—a person or an AI system—fails to register a morally relevant feature of the world or fails to register its moral importance. When users rely on a system that lacks moral discernment it threatens the user's autonomy by making them dependent on a process for decision making that does not incorporate a relevant ethical concern. It threatens the user's welfare because they are dependent on a process that lacks information necessary to act in ways that safeguard their interests or advance their goals and plans.

Concerns about moral discernment arise for SAs that seek to provide proactive assistance because such systems take on cognitive tasks that are currently reserved for the human user (or persons in their social support network). But moral discernment is a challenge for AI systems because of the diversity in, and lack of uniformity among, features of social interactions that reliably indicate morally relevant events. Morally relevant facts often supervene on or are realized by a wide range of physical states of the world. Deciphering the meaning of social interactions is particularly difficult because that meaning is determined by context and information relevant to determining context can be distributed over a long series of prior interactions [52], [53]. Similarly, morally relevant features of an interaction often derive, not from the literal content of what a person says, but from the speech acts that the agent uses those utterances to perform. Speech acts are actions that individuals perform in conversation with others [54], [55]. A promise, for example, is a speech act through which A commits to doing something for B. Other speech acts include apologizing, soliciting an offer of some kind, making an offer, and commanding.

To illustrate these concerns, consider the following hypothetical interaction between Ann, a woman in her seventies, and her primary care physician (PCP).

Ann is at a regular checkup when her PCP expresses concern that she is showing signs of type-2 diabetes. After asking a range of questions her PCP says, "I'm afraid you know what I'm going to recommend."

"You're like a broken record," Ann replies.

"Yes, and now it looks like we may be seeing some of the results that I've been warning you about. Tingling in your feet, your dry skin, and increased thirst, these are all part of a cluster of warning signs that you are becoming diabetic. We need to run some tests and when your labs come back, if your sugar is elevated, you will need to see an eye care specialist. Diabetes-related vision loss is a major concern, and it's been some time since you've had an eye exam."

"Ok, I don't mind seeing a specialist for my eyes."

"Excellent. And we have talked a lot about those sweets. What do you say? I'm sure we can find delicious alternatives that are better for your blood sugar."

Ann looks at her PCP and after a moment of silence says, "Diabetes. I guess I should not be surprised. If you think it will help, what do I have to lose?"

"Excellent!" her PCP says, "Trust me, you will have more energy and it will be better for your knees if you slim down even a little. Baby steps are fine as long as we're moving in the right direction."

This exchange is complex and multi-layered in terms of its informational content: Ann learns that she is at risk of diabetes and of diabetes-related vision loss. She learns that her physician will order blood work and that if it comes back with certain results, she will need to see a specialist. In this exchange, an important speech act has also been performed—Ann has committed herself to seeing both an eye care specialist and a dietitian. Although eye examination is mentioned explicitly, unspoken in this exchange is the fact that, for several prior meetings, Ann's PCP has been urging her to reduce her sugar intake and to consult with a dietician to help arrange a menu of meals and snacks that Ann will still enjoy while allowing her to reduce her caloric and sugar intake. Ann's PCP brings that open question about whether she will see a dietitian back to the forefront of the conversation when she brings up sweets and the possibility of finding healthier alternatives. Ann's PCP can gauge Ann's understanding of the open question and takes her statement "What do I have to lose?" as willingness to give the dietitian a try. But there is no guarantee that a third party who was not privy to these prior conversations would know that this question has been revisited, let alone that a proposal made at a prior meeting has been accepted.

Remembering each of the pieces of information conveyed in this exchange might be a challenge for any patient. But it is likely to be especially challenging for an older adult grappling with declining cognitive abilities. If Ann were accompanied by a human companion, that person would likely have little trouble making a succinct record of the events in this interaction. If that person were unaware of the content of Ann's past interactions with her PCP, then they would likely understand that a prior office visit is being referenced and inquire about the nature of the agreement that had been made.

To fulfill a similar role to a human caregiver, an SA would have to situate this conversation in its larger relational context. Human conversations are governed by pragmatic norms that aim to make discourse cooperative and efficient [55]. Ann and her PCP re-open an ongoing conversation about Ann's dietary needs without explicitly stating the question under discussion. Rather, their shared understanding that new information is relevant to this open question allows them to resume a previous, salient conversation expeditiously. In particular, when Ann almost-rhetorically asks, "What do I have to lose?" she is not asking a question, she is signaling a commitment to act—to undertake a new course of action designed to help her advance her health goals. Her words are important, not so much for their literal content, but because they constitute the acceptance of a commitment. Without a model of the shared conversational history between these speakers, Ann's SA would not be able to situate their discourse in its proper context. Without that ability, the SA would be oblivious to this aspect of Ann's social interaction with her PCP and this lack of moral discernment would jeopardize Ann's autonomy and wellbeing.

The details presented in the example are not as important as the more general points the scenario is intended to illustrate, namely that: 1) information about the context in which humans interact can be spread out over time and place, 2) speech acts are a ubiquitous aspect of human social interactions in which the words that a speaker utters are not a direct guide to understanding the action that the speaker has performed in uttering those words and 3) that speech acts are often the means through which agents perform actions with important moral content or implications. The ability of an assistive system to identify ethically relevant speech acts is thus critical if it is to assist with social interactions effectively. Although some speech acts use tokens that are easy to identify, these morally significant acts can be accomplished through a wide variety of utterances or behaviors. That is, although acts of promising can involve a canonical identifying utterance, such as, "I promise," this need not be the case. Uncountably many different utterances, and even non-verbal cues, can be used to make a promise. In that regard, an SA that listens in on a conversation but that cannot apprehend the non-verbal behaviors of speakers is likely to be insensitive to morally relevant information that is communicated through a medium to which it does not have access. The resumption of questions under discussion and the complex way speech acts can be performed are just two examples of the way that morally relevant information can depend on context and go beyond the kind of concrete features of the world that current AI systems are adept at detecting.

Promises are only one type of speech act; many other types of speech acts may pose significant challenges for an SA [56]. Yet, we focus on promises here because they have special moral importance as a ubiquitous and meaningful way that agents take on obligations and transfer entitlements to others. If A has made a promise to B, then B not only has a legitimate expectation that A will perform the promised action but an entitlement to A's performance. If all other things are equal, A's promise creates an obligation on A to carry out the promised act and satisfy this entitlement to B. Promises can differ in their moral force or importance depending on their role in the life of the parties involved or others who might be affected. Missing a ride to a child's birthday party might be a welcome relief for a busy family member whose absence is

unlikely to spoil the occasion. But missing a ride to the same party for a grandparent who delights in any interaction with grandchildren could be a significant diminution of welfare. This illustrates the concern that older adults experiencing cognitive decline are vulnerable to a range of physical, psychological, social, or economic harms from failures on the part of a system to track and record information that they rely on to discharge their commitments and obligations.

Summarizing the content of informationally rich social interactions is an example of a way future SAs might provide proactive support in social interactions. But an SA that is only capable of summarizing the literal contents of what is asserted in a verbal interaction will be oblivious to speech acts that are accomplished through those utterances, much less through non-verbal cues. Moreover, systems that cannot situate a social interaction into a longer history of prior interactions may be oblivious to key contextual features such as open questions, expectations, and prior commitments. Individuals who rely on such systems would thus face potentially serious risks to important moral interests unless the division of cognitive labor here could be rebalanced in a way that would mitigate these risks without undermining the utility of the system by shifting the relevant cognitive load back onto the user. We return to this issue in section VI.

## V. The Structure of Commitments and Conflicting Obligations

Another area in which advanced SAs may provide more proactive assistance to older adults involves scheduling. Although not normally thought of in such grandiose terms, calendars and schedules are tools that people use to represent and track their obligations and entitlements. Managing a calendar or schedule requires some knowledge of the relative importance and relationship among the web of obligations and entitlements that it represents. In the previous section, we saw how a lack of moral awareness can impact the autonomy and welfare of older adults to the extent that it deprives them of information they need to advance their goals and to live up to their commitments and obligations. Even when systems are aware of the user's obligations and entitlements, systems that attempt to assist with scheduling are taking on cognitive tasks that can make the user vulnerable to conflicting obligations and frustrated entitlements.

To illustrate this idea, imagine that Ann is accompanied to her appointment by a human caregiver who creates a schedule from the exchange mentioned in the previous section. Part of that schedule involves an appointment to give blood and then waiting to hear whether Ann needs to see an eye specialist. If Ann does not hear about those results in a few days, her human caregiver might check in with the PCP's office. Likewise, imagine that the results come back as expected and Ann must visit an eye specialist. Her caregiver arranges an appointment for next Thursday at 10:00am at an office downtown. Because Ann's daughter lives downtown, Ann arranges to meet her for breakfast on the day of her appointment. Since Ann no longer drives, she or her caregiver must create a timetable for leaving her house on the day of the appointment and commuting downtown via rapid transit. However, three days before the appointment Ann receives a call from her PCP's office, reminding her of the appointment and informing her that she must fast starting 12 hours before the appointment. Ann's caregiver would likely recognize the conflict between the scheduled breakfast and the requirements for Ann's PCP appointment. In that case, Ann keeps the morning appointment with the specialist but reschedules the meeting with her daughter to lunchtime.

Although scheduling is a routine activity, it requires complex cognitive functions that draw on extensive, and difficult to represent, background knowledge. This includes knowledge about what is required to discharge a commitment, an understanding of the importance of a commitment, the stakeholders who bear relevant responsibilities with respect to a commitment, the conditions under which commitments conflict, and strategies for mitigating or resolving those conflicts given their relative importance. For example, although working out a contingency plan for every event on a calendar might be unnecessary, Ann's ability to safeguard her vision hinges on a division of labor in which the laboratory is expected to share the results with her PCP's office, which then is responsible for contacting Ann. If the results are not reported back in a certain amount of time, then one of these parties may have fumbled their responsibility. The importance of this step, within a scenario in which a person's vision might be at risk, warrants being vigilant about such contingencies.

Similarly, when Ann schedules an appointment with her PCP, she takes on a series of commitments including a defeasible obligation to be at a certain place at a certain time and in a certain state (to have fasted for twelve hours before her appointment). This obligation is defeasible in the sense that it can be overridden if something more important arises. But overriding an obligation can create problems—delays in testing might expose Ann to health risks or her provider may charge a fee for appointments that are not canceled by a certain date. If no such overriding obligation arises, Ann's obligation to be at the appointment constrains her liberty by limiting her from accepting other conflicting commitments. Thus, a breakfast meeting with her daughter is inconsistent with the commitments that are in force for Ann on that day.

While a current SA could detect when a schedule is double booked, realizing that Ann cannot meet her daughter for breakfast requires the ability to understand a conflict in commitments in a dimension other than a literal overlap in time. Many individual events are points in a larger series of interactions that entail a wide range of requirements. Many social gatherings, from birthday parties and weddings to more informal meetings can entail purchasing a new outfit, buying a gift, or in some sense not showing up "empty-handed" or in the wrong state (not having fasted before a medical test). But not showing up empty-handed, or in the right attire, or in the right condition requires a series of activities that must themselves be accomplished prior to the event in question. Similarly, some travel destinations require reservations in advance or some type of authorization, such having a visa, a satisfactory vaccination record, or a valid passport. In this sense, travel often involves more than a series of dates for departure and arrival since many steps in the travel process require extensive advance preparation. Attending an event or traveling can require

understanding the norms in force, who the relevant stakeholders are, and what it takes to satisfy these requirements.

Hence, the entries on calendars and schedules are points in time intimately connected to implicit and complex strategies for carrying out plans, meeting our obligations, and exercising our entitlements. Humans might overlook this complexity since we draw on relevant background knowledge so easily to navigate this complex, moral, action space. For future SAs to take a more proactive role in scheduling—shouldering some of the cognitive load—they would have to detect and represent the broader plans, relationships, goals, normative expectations, and obligations of the person for whom they are scheduling including how to reconcile potentially conflicting demands.

Avoiding intractable conflict may also require the system to represent who the relevant duty bearers are in tasks for which the user is only one party in a larger division of labor (as when Ann awaits the results from her test before she can make an appointment with her specialist), to apprehend the relevant time horizon for various responsibilities (e.g., when to suspect something has gone wrong with the test and what the time horizon is for visiting a specialist), to represent which actions or states represent a discharge of that duty and which represent a failure to discharge (e.g., receiving the test or not hearing anything for a specific number of days), and to formulate an alternative plan if other agents fail to discharge their responsibilities (e.g., calling the PCP if test results are not back within a certain timeframe). If older adults rely on an advanced SA to keep their schedule, but it lacks some of the background knowledge necessary to identify conflicting commitments or to identify additional steps necessary to satisfy norms that are in force for an event, then users might find themselves in situations in which they cannot satisfy one or more of their commitments or in which they cannot take advantage of a series of opportunities that might have been open to them with more intelligent planning.

## VI. Teaming and Net Benefit

In the previous sections, we focused on the way that advanced SAs designed to be proactive in their assistance could create certain opportunities and vulnerabilities for users. The opportunities involve offloading some cognitive tasks to the SA and taking advantage of its ability to assist—to help the user map out, schedule, and execute their plans. The vulnerabilities stem from the complexities of the cognitive operations involved in these tasks and from the way that a system's failure to perceive morally relevant information could adversely impact the interests of the user. These adverse impacts include frustrating the user's ability to take advantage of entitlements, impeding user autonomy, impairing user wellbeing or generating conflicting obligations that a user cannot jointly satisfy.

One strategy to mitigate these vulnerabilities is for the system to interact with the user in ways that increase the likelihood that it has an accurate representation of morally relevant information. This includes the ethical significance of events that have transpired, the various projects the user is trying to accomplish and their relative importance, the content of the obligations the user has accepted, the entitlements the user wants to exercise, and the norms with which the system needs to comply to achieve the user's goals. To accomplish these goals, a future SA might engage the user in dialogue about these issues, seeking user feedback to validate the system's representation of key information and identify and address relevant shortcomings. Dialogue might also provide the user with the opportunity to understand decisions the system has made in order to provide oversight in a way that seems natural and engaging to the user.

However, facilitating effective teaming between users and SAs poses its own challenges. The utility of advanced SAs is supposed to derive from their ability to reduce the cognitive load on users while helping them to accomplish tasks that are important to their ability to function. Monitoring for failures of moral discernment without increasing cognitive load on users would require SAs to engage in a complex, higher-order task of evaluating the completeness of their own representation of the various ethical issues just mentioned. Systems that suffer from a first-order failure of moral discernment (e.g., a system that fails to record that the user took on a moral commitment by making a promise) are unlikely to be more discerning at a higher-order level. Rather, the same lack of discernment is likely to propagate to higher levels in the system. Conversely, systems that are fastidious about engaging users in conversations that audit their first-order representations of key ethical issues are likely to increase the cognitive load for users. At one extreme, a system that is constantly asking the user whether events are salient, what the content of an obligation or entitlement is, and so on, risks becoming overly intrusive and requiring so much time and attention that the user no longer enjoys a net reduction in time and effort from using the system.

More generally, for the system to accomplish the goal of providing assistance it must produce a net benefit while requiring less effort and attention from the user than what they would have expended in its absence. If Ann must spend a lot of time monitoring the SA to ensure that it performs reliably, or if the oversight process is itself so complex or challenging that Ann has difficulty successfully carrying out required steps, then the system may no longer be a convenience. It only succeeds in providing assistance to the extent that it enables the user to accomplish tasks with less effort and frustration than if the user tried to accomplish those tasks on their own or through the next best alternative.

Other responses to concerns about failures of moral discernment face similar problems. For example, a future SA might save a recording or a transcript of Ann's conversation with her PCP. If it fails to identify that Ann took on a commitment to see a dietician, perhaps Ann could recover that information from the recording or the transcript. But if the system cannot register that a morally salient event took place how would it know to prompt Ann to review the transcript of the meeting? Relying on Ann to know to do this offloads onto her the more complex, higher-order task of monitoring her SA for failures of moral discernment. But this requires Ann to exercise the very capacity with which the system is intended to provide support. It also presumes that the user's ability to focus and to extract the relevant information will be superior on the second go round. This might be the case, since the ability to pause a recording might help the user to identify and extract different elements from an informationally dense exchange.

But it also depends on the user's ability to sustain focus on the extractive task throughout the duration of the recording, while checking on the accuracy of the SA's summary of the event—a task that might itself be challenging for some users.

A second approach is to educate users about the SA's capabilities and limits. This might involve ensuring that users only use the SA for tasks it is capable of performing reliably. For example, if the SA is incapable of identifying some class of speech acts, then alerting the user to this might enable them to anticipate areas in which the system cannot perceive morally relevant information. However, understanding an SA's limitations may be insufficient to mitigate its deficiencies if it requires an older adult, experiencing the initial stages of cognitive decline, to perform a complex second-order task on top of the complexities of the first-order social interaction. In other words, Ann might not be able to attend to the first-order task of interacting with her PCP during an information-dense exchange and to carry out the second-order task of monitoring the conversation for speech acts that her SA might not be able to identify.

Like with autonomous vehicles, the benefits of proactive assistive systems may not increase proportionally with the degree of assistance that a system provides to the user. Until the assistive system reaches a high threshold of reliability, it may be preferable for the system to rely on the user to perform complex cognitive tasks. It seems reasonable for this threshold to vary with the risk or importance of the task. But, given the ambition of this type of system to be proactive, the question arises as to how the user will ensure that they only rely on the system, or that it only undertakes tasks, that fall below the relevant risk threshold.

## VII. Assistance or Influence

To be assistive, a system must enhance the ability of the target of assistance to pursue and effectuate projects, plans, and activities in which they find meaning and fulfillment. Recollecting information and maintaining a schedule are examples of all-purpose cognitive tasks in the sense that these tasks could be required by a wide range of social activities. However, when users rely on systems for support, they become vulnerable to the limitations of those systems. Limitations in the ability of an SA to carry out an all-purpose cognitive task can have a profound impact on users by altering the mix of activities in which they engage. The reason is that the system's capacity to support the relevant task in a particular domain, or with respect to a particular subject matter or stakeholder make it easier for the user to carry out those tasks than in other domains, involving different subjects or stakeholders. This kind of influence can happen inadvertently when, for example, interactions in some environments are just more difficult for the system to manage, such as trying to identify distinct conversations in a crowded party versus tracking the slow and steady speech of a professional making a special effort to be understood. But this kind of influence could also be intentional if systems are trained or designed to encourage users to engage in specific activities and to ignore others.

Whether a system succeeds in providing assistance or exerts some other form of influence on the user depends, in part, on the extent to which the activities it supports are central to the user's wishes. When supported activities fall at the periphery of the user's projects and plans, then shifting their activity mix can unintentionally frustrate user autonomy and welfare. This can occur by making the activities users value most highly more difficult or more costly to undertake. To the extent that these costs encourage users to forego those activities in favor of less valued but more easily supported activities, this can result in a diminution of user wellbeing. In these cases, the dependence of the user on a system that cannot support the projects, plans or activities most valued by the user shifts from a supportive relationship to a relationship of unintended but undue influence.

One remedy for the vulnerability of older adults to the unintended influence of limitations from SAs is to periodically audit the degree of fit between the goals, projects and activities that users value and what to pursue, the extent to which they rely on the SA for assistance with these activities, and the extent to which the SA provides needed support. When SAs fall short, providing users with alternative supportive assistance might enhance their ability to pursue those activities they most value and from which they find the most fulfillment. However, this process can be challenging in cases where the SA is intended to supplant other supportive services rather than to augment them.

Finally, safeguards must be implemented to ensure that future SAs are not a means by which a third party can exploit users for its own purposes. For example, a user who relies on a SA to order groceries might be manipulated if the system is designed only to order products from companies that have a business relationship with the device's manufacturer. While such a relationship might be profitable for the sponsor and the manufacturer, restricting the offerings to those preferred by sponsors might influence dietary choices in ways that affect user enjoyment and health.

## VIII. Assistance or Commodification

The ambition of developing future SAs that take advantage of a wider range of multi-modal sensing and that keep track of a long history of sensitive information across multiple social and personal domains compounds and amplifies privacy concerns. SAs that seek to provide proactive assistance in a social space will require intimate knowledge not only of the user's projects, plans, values, and physical and mental health status, but of their commitments and relationships with others and how they value those relationships. Modeling this information will involve longitudinal monitoring of intimate and highly sensitive personal information. For this information to be useful for the system itself, it cannot be anonymized in real time; longitudinal data must be correlated with the individual to monitor health trends. To provide proactive assistance an AI will likely need to make predictions about the rate of decline in the user's abilities, generating a sensitive piece of medical information to which the user may or may not want access and to which the user likely would not want others to have access.

Ensuring that a future SA facilitates social interactions in a way that respects user privacy replicates some of the problems already discussed. Sharing information in ways that track Ann's preferences and the permissions and entitlements in force in her social relationships is an exercise in moral discernment. Oversharing information makes Ann vulnerable to violations of her privacy and confidentiality. Undersharing information may

deprive caregivers, family or friends of the information they need to respond to Ann's needs. Here too, this task is complicated by the dynamic nature of norms and the importance of a wide range of counterfactuals. For example, Ann may be comfortable with the SA sharing medical information with her daughter but not other relatives. However, if Ann's mobility decreases before an appointment and Ann's daughter is not available for assistance, how much information should the SA disclose to other relatives to secure their help? There is also the potential for conflict here: how should the system navigate a situation in which it is tasked with facilitating Ann's doctor appointment, but Ann has forbidden sharing for certain types of information (e.g., that Ann smoked a cigarette recently)? These questions are likely to be common in the context of a user's declining mental capacity.

Privacy issues also arise for future SA's at a different level. Current SAs are used to generate data that companies use for business purposes. Given the sensitive and ubiquitous nature of the data collected by the kind of near-future SA that we envision here, it is unclear that such a model would be ethically permissible even if it were practically feasible. It might not be feasible since users who understand the extent to which business would have access to sensitive and intimate information about some of the most sensitive aspects of their lives might not want to use such a system. But like many complex systems, most users might not fully understand the extent to which such systems generate private, identifiable information about intimate and sensitive aspects of their lives.

More fundamentally, however, relationships of assistance that involve sensitive aspects of a person's agency, autonomy or wellbeing entail fiduciary obligations on the part of those who provide assistance. Fiduciary obligations are obligations to give primacy to the interests of the party receiving assistance relative to the interests of the party providing assistance. These fiduciary obligations are grounded in the nature of the assistance that is offered, the degree to which the person who relies on this assistance is made vulnerable to harm or wrongdoing from this relationship of reliance or dependence, and the infeasibility of the target of assistance being able to monitor and provide effective oversight over the party providing assistance. Health care relationships are regarded as fiduciary in nature because they meet all of these criteria: a person's health status involves sensitive private information and implicates the ability of a person to function in ways that they value, individuals are vulnerable to a wide range of harms to their agency or wellbeing when they submit to medical care and it is impossible for lay people to provide effective oversight of health care services since those services frequently require specialized knowledge, skills or abilities that non-medical professionals lack.

The privacy issues raised by future SAs thus pose daunting technical, social, and ethical challenges. These challenges arise at the level of designing a system capable of providing effective assistance that merits user trust and reliance. They also arise at the level of the business model used to support the development of such a system. To be successful, this business model must be capable of sustaining investment in the ecosystem necessary to support and maintain these systems while respecting the deeply private and sensitive nature of the information such systems will require to fulfill their stated function.

## IX. Vulnerability to Injustice

The distinct ethical concerns we have raised illustrate how reliance on AI systems creates vulnerabilities that implicate the wellbeing and the autonomy of older adults. Because these concerns flow from the three interrelated dynamics mentioned previously, they are likely to be closely connected in practice. Understanding these interrelationships illustrates how reliance on AI systems can render this population vulnerable to injustice.

Justice requires that moral equals be treated equally, including meaningful social recognition of the interest of every person in having real freedom to formulate, pursue and revise an individual life plan [57]. When groups face systematic and avoidable social dynamics that undermine this freedom, they have a credible claim to experiencing injustice. As we age, we frequently require assistance with tasks necessary to revise and pursue our individual life plans. An important component of such assistance involves understanding the way that changes in a person's capabilities and life circumstances alter their ability to advance life projects and how their deeply held goals and values might find new outlet through relationships or activities that they have to ability to pursue. In person-to-person relationships, this often requires empathy and credible effort to understand a person's perspective on the world. When groups are systematically denied the empathy and effort required to bring credible attention to their interests, they suffer from what has been called hermeneutical or epistemic injustice [58].

Without solutions that are sensitive to the dynamics we have outlined, older adults who rely on SAs are likely to be subject to hermeneutical injustice and potentially other justice-related harms. The systems on which they rely to support their capacity to function will be unable to comprehend some aspect of their interests, how their interests might adapt to their changing capabilities and circumstances, or be unable to identify ethically relevant social acts or relationships that affect their interests. They are also likely to be subject to manipulation—forms of influence that advance ends that are not their own—and to intrusive information gathering and commodification. Together, these problems constitute a form of social domination—a vulnerability to more powerful parties arbitrarily interfering with and subverting one's freedom to form, pursue, and revise a meaningful life plan [59].

As stakeholders make progress on the issue, they must also take care to ensure that systems function safely and effectively for older adults from different backgrounds or abilities, including those with different accents or speech impairments due to illness or disability.

## X. Conclusion

The ambition of using AI to provide support for older adults holds out significant promise as a socially and individually valuable application of near-future AI technology. The purpose of this paper is to anticipate novel ethical challenges that arise from developing systems with two properties likely necessary to achieve this goal—the ability to proactively anticipate user

needs and the ability to provide assistance, at least in part, by mediating social relationships. These novel ethical challenges derive from several factors. Providing proactive assistance with even seemingly simple tasks involving memory and scheduling involve extracting morally relevant information that is highly context dependent, whose content derives partly from a rich background set of norms and expectations and that can supervene on, or be realized by, a wide range of utterances and nonverbal acts. Relying on such systems creates special vulnerabilities for user autonomy, wellbeing, privacy, ability to effectively manage their social relationships and commitments, and to maintain control over valuable life projects. Mitigating these risks faces challenges stemming from the cognitive limitations of users, the complexity of monitoring systems that seek to be proactive, and the prospect that overly demanding monitoring requirements can undermine the utility of assistive systems.

Our hope is that articulating these dynamics and their moral importance will make these challenges salient to developers and encourage them to be transparent in terms of how incremental advances in AI technology might exacerbate, mitigate, or resolve some of these tensions. Providing credible assurance that assistive systems are reliable and robust relative to these challenges is likely a critical step to warranting trust from users and to ensuring that assistive systems provide a net benefit to users, rather than merely redistributing the way they spend their time and energy.

## REFERENCES

[1] Masina, F., Orso, V., Pluchino, P., Dainese, G., Volpato, S., Nelini, C., … Gamberini, L.,. "Investigating the accessibility of voice assistants with impaired users: Mixed methods study," *Journal of Medical Internet Research*, vol. *22*(9), 2020, doi: 10.2196/18431.

[2] D Graham, S. A., Lee, E. E., Jeste, D. V, Van Patten, R., Twamley, E. W., Nebeker, C., … Depp, C. A., "Artificial intelligence approaches to predicting and detecting cognitive decline in older adults: A conceptual review,"*PsychiatryResearch*, pp*284*, Art. no.112732, 2020, doi: 10.1016/j.psychres.2019.112732

[3] Howe, W., andYampolskiy, R., "Impossibility of Unambiguous Communication as a Source of Failure in AI Systems," in *Proc. AISafety@ IJCAI, Montreal, CA, 2021*.

[4] Wangmo, T., Lipps, M., Kressig, R. W., and Ienca, M., "Ethical concerns with the use of intelligent assistive technology: Findings from a qualitative study with professional stakeholders," *BMC Medical Ethics*, vol. *20*(1) 2019, 1–11. doi: 10.1186/s12910-019-0437-z.

[5] Mäyrä, F.,and Vadén, T., "Ethics of living technology: Design principles for proactive home environments," *Human IT*, vol. *7*(2), pp. 171–196, 2004.

[6] Ogonji, M. M., Okeyo, G., Wafula, J. M., "A survey on privacy and security of Internet of Things," *Computer Science Review*, Art. no. *38*, Art. no. 100312, 2020,doi: 10.1016/j.cosrev.2020.100312

[7] Belk, R., Ethical issues in service robotics and artificial intelligence. *Service Industries Journal*, vol. *41*(13–14), pp. 860–876, 2021, doi: 10.1080/02642069.2020.1727892

[8] Elahi, H., Wang, G., Peng, T., and Chen, J., "On transparency and accountability of smart assistants in smart cities," *Applied Sciences (Switzerland)*, vol. *9*(24), 2019, doi: 10.3390/app9245344

[9] Sharkey, A., and Sharkey, N., "Granny and the robots: Ethical issues in robot care for the elderly," *Ethics and Information Technology*, vol. *14*(1),pp. 27–40, 2012, doi: 10.1007/s10676-010-9234-6

[10] J. Miller, T. McDaniel and M. J. Bernstein, "Aging in Smart Environments for Independence," IEEE International Symposium on Technology and Society (ISTAS), 2020, pp. 115-123, doi: 10.1109/ISTAS50296.2020.9462211.

[11]Miller, Jordan, and Troy McDaniel. "Socially Assistive Robots for Storytelling and Other Activities to Support Aging in Place," In Multimedia for Accessible Human Computer Interfaces, pp. 145-172. Springer, Cham, 2021, doi: 10.1007/978-3-030-70716-3_6.

[12]Shahriari, Kyarash and Shahriari, Mana, "IEEE standard review — Ethically aligned design: A vision for prioritizing human wellbeing with artificial intelligence and autonomous systems," pp. 197-201, 2017, doi: 10.1109/IHTC.2017.8058187.

[13] D. Schiff, J. Borenstein, J. Biddle and K. Laas, "AI Ethics in the Public, Private, and NGO Sectors: A Review of a Global Document Collection," *IEEE Transactions on Technology and Society*, vol. 2, Art. no. 1, pp. 31-42, March 2021, doi: 10.1109/TTS.2021.3052127.

[14]Defense Innovation Board, "AI Principles: Recommendations on the Ethical Use of Artificial Intelligence," pp. 1–74, 2019.

[15]Hagendorff, T, "The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds and Machines*," vol. *30*(1), pp. 99–120, 2020, doi:10.1007/s11023-020-09517-8.

[16]Shahriari, K., and Shahriari, M., "IEEE standard review — Ethically aligned design: A vision for prioritizing human wellbeing with artificial intelligence and autonomous systems," in *IEEE Canada International Humanitarian Technology Conference (IHTC).* pp. 197–201,2017, doi:10.1109/IHTC.2017.8058187

[17]University of Montreal, "Montreal Declaration Responsible AI," 2018. [Online]. Available:https://www.montrealdeclaration-responsibleai.com/the-declaration

[18]Merritt, S. M., and St, M., "Affective Processes in Human – Automation Interactions," 2004, doi: 10.1177/0018720811411912

[19]Boada, J. P., Maestre, B. R., and Genís, C. T., "The ethical issues of social assistive robotics: A critical literature review," *Technology in Society*, vol.*67, Art. no. 101726, 2021, doi: 10.1016/j.techsoc.2021.101726*.

[20]Vandemeulebroucke, T., Dierckx de Casterlé, B., and Gastmans, C., "The use of care robots in aged care: A systematic review of argument-based ethics literature," *Archives of Gerontology and Geriatrics*, vol *74*pp. 15–25, September 2017,doi: 10.1016/j.archger.2017.08.014.

[21]Venter, S. L., Olivier, M. S., and Britz, J. J. , "Toward a model of responsibility for proactive systems," *Journal of Information Ethics*, vol. *17*(2), pp. 78–90, 2008, doi: 10.3172/JIE.17.2.78.

[22]Woods, D. D., "Cognitive technologies: The design of joint human-machine cognitive systems," *AI Magazine*, vol. *6*(4), pp. 86, 1985, doi: 10.1609/aimag.v6i4.511.

[23]Kuusela, K., Koskinen, I., and Mäyrä, F., "A metamorphosis of the home: Proactive information technology as a design challenge," *Nordes 2005: In the Making*, vol. *24*, 2005, doi: 10.21606/nordes.2005.022.

[24]Danaher, J., "Toward an Ethics of AI Assistants : an Initial Framework,"*Springer Nature*, vol. 31, pp. 629–653, December 2018, doi: 10.1007/s13347-018-0317-3.

[25]Castelfranchi, C., and Falcone, R., *Trust Theory: A socio-cognitive and computational model*,vol. 18, West Sussex: John Wiley & Sons, May 2010.

[26]Desai, M. U., and Yanco, H.,, "Modeling Trust To Improve Human-Robot Interaction," 2012.

[27]Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y. C., de Visser, E. J., and Parasuraman, R., "A Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. *53*(5), pp. 517–527, doi:10.1177/0018720811417254.

[28]Wagner, A. R.,"The role of trust and relationships in human-robot social interaction," Ph.D. dissertation, College of Computer Science, Georgia Institute of Technology, Atlanta, GA, December 2009.

[29]Komiak, S. Y. X., and Benbasat, I., "The effects of personalization and familiarity on trust and adoption of recommendation agents," *MIS Quarterly: Management Information Systems*, vol. *30*(4), pp. 941–960, 2006, doi: 10.2307/25148760.

[30]Miller, C. A., & Parasuraman, R., "Designing for flexible interaction between humans and automation: Delegation interfaces for supervisory control," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. *49*(1), pp. 57–75, 2007, doi: 10.1518/001872007779598037. PMID: 17315844.

[31]Razin, Y. S.,"Interdependent Trust for Humans and Automation Survey," 2020, [Online]. Available: http://www.cognitiveengineering.gatech.edu/presentations/new-interdependent-trust-humans-and-automation-i-thau-survey

[32]Gefen, D., Karahanna, E., and Straub, D. W.,"Trust and TAM in Online Shopping: An Intergrated Model," *MIS Quaterly*, vol. *27*(1), pp. 51–90, 2003, doi: 10.1017/CBO9781107415324.004.

[33] McKnight, D, H., Carter, M., Thatcher, J. B., and Clay, P. F., "Trust in a Specific Technology : An Investigation of Its Components," vol. *2*(2), 2011 doi: 10.1145/1985347.1985353

[34] McKnight, D. H., and Chervany, N. L.,"What trust means in e-commerce customer relationships: An interdisciplinary conceptual typology," *International Journal of Electronic Commerce*, vol. *6*(2),pp. 35–59, 2001, doi:10.1080/10864415.2001.11044235

[35] Lee, J. D., and See, K. a., "Trust in automation: designing for appropriate reliance," *Human Factors*, vol. *46*(1), pp. 50–80, 2004, doi: 10.1518/hfes.46.1.50.30392.

[36] Robinette, P., Howard, A., and Wagner, A. R., "Conceptualizing overtrust in robots: why do people trust a robot that previously failed?," *Autonomy and Artificial Intelligence: A Threat or Savior?,* pp. 129–155, Springer, 2017, doi: 10.1007/978-3-319-59719-5_6.

[37] Merritt, S. M., Unnerstall, J. L., Lee, D., and Huber, K., "Measuring Individual Differences in the Perfect Automation Schema," *Human Factors*, vol. *57*(5), pp. 740–753, 2015, doi: 10.1177/0018720815581247.

[38] Wang, R., Harper, F. M., and Zhu, H., "Factors influencing perceived fairness in algorithmic decision-making: Algorithm outcomes, development procedures, and individual differences," in *Proc.CHI Conference on Human Factors in Computing Systems,* Honolulu, HI, USA, 2020, pp. 1–14.

[39] Booth, S., Tompkin, J., Pfister, H., Waldo, J., Gajos, K., and Nagpal, R., "Piggybacking robots: Human-robot overtrust in university dormitory security," in *Proc.ACM/IEEE International Conference on Human-Robot Interaction,* Vienna, AT, 2017, pp. 426–434.

[40] De Siqueira, A. S. S., Yokomizo, J. E., Jacob-Filho, W., Yassuda, M. S., and Aprahamian, I., "Review of Decision-Making in Game Tasks in Elderly Participants with Alzheimer Disease and Mild Cognitive Impairment," *Dementia and Geriatric Cognitive Disorders*, vol. *43*(1–2), pp. 81–88, 2017, doi: 10.1159/000455120.

[41] Landau, R., a,d Werner, S., "Ethical aspects of using GPS for tracking people with dementia: Recommendations for practice," *International Psychogeriatrics*, vol. *24*(3), pp. 358–366, 2012, doi: 10.1017/S1041610211001888.

[42] Borenstein, J., & Arkin, R., "Robotic Nudges: The Ethics of Engineering a More Socially Just Human Being," *Science and Engineering Ethics*, vol. *22*(1), pp. 31–46, doi: 10.1007/s11948-015-9636-2.

[43] Capasso, M., & Umbrello, S. (2021). Responsible nudging for social good: new healthcare skills for AI-driven digital personal assistants. *Medicine, Health Care and Philosophy*, (0123456789). https://doi.org/10.1007/s11019-021-10062-z.

[44] Berdasco, A., López, G., Diaz, I., Quesada, L., and Guerrero, L. A., "User experience comparison of intelligent personal assistants: Alexa, Google Assistant, Siri and Cortana," *Multidisciplinary Digital Publishing Institute Proceedings*, vol. *31*(1), pp. 51, 2019, doi: 10.1016/j.techsoc.2021.101726.

[45] Lin, F. R., "Hearing loss and cognition among older adults in the United States," *Journals of Gerontology - Series A Biological Sciences and Medical Sciences*, vol. *66 A*(10), pp. 1131–1136, 2011 doi: .1093/gerona/glr115.

[46] Monge, Z. A., and Madden, D. J., "Linking cognitive and visual perceptual decline in healthy aging: The information degradation hypothesis," *Neuroscience and Biobehavioral Reviews*, vol. *69*, pp. 166–173, 2016, doi: 10.1016/j.neubiorev.2016.07.031.

[47] Yakhno, N. N., Zakharov, V. V., and Lokshina, A. B., "Impairment of memory and attention in the elderly," *Neuroscience and Behavioral Physiology*, vol. *37*(3), pp. 203–208, 2007, doi: 10.1007/s11055-007-0002-y.

[48] Kocaballi, A Baki, Ijaz, K., Laranjo, L., Quiroz, J. C., Rezazadegan, D., Tong, H. L., … Coiera, E., "Envisioning an artificial intelligence documentation assistant for future primary care consultations: A co-design study with general practitioners," *Journal of the American Medical Informatics Association*, vol. *27*(11), pp. 1695–1704, doi: 10.1093/jamia/ocaa131.

[49] Lacson, R. C., Barzilay, R., and Long, W. J., "Automatic analysis of medical dialogue in the home hemodialysis domain: structure induction and summarization," *Journal of Biomedical Informatics*, vol. *39*(5), pp. 541–555, 2006, doi: 10.1016/j.jbi.2005.12.009.

[50] Quiroz, J. C., Laranjo, L., Kocaballi, A. B., Berkovsky, S., Rezazadegan, D., and Coiera, E., "Challenges of developing a digital scribe to reduce clinical documentation burden," *NPJ Digital Medicine*, vol *2*(1), pp. 1–6, 2019, doi: 10.1038/s41746-019-0190-1.

[51] Clark, A., and Chalmers, D.,. "The extended mind," *Analysis*, vol. *58*(1), pp. 7-19, 2019, doi: 10.1093/analys/58.1.7.

[52] Chen, M. X., Lee, B. N., Bansal, G., Cao, Y., Zhang, S., Lu, J., … others., "Gmail smart compose: Real-time assisted writing," in *Proc. 25th ACM SIGKDD International Conference on Knowledge Discovery \& Data Mining*, Anchorage, AK , USA, 2019 pp. 2287–2295.

[53] Kocaballi, Ahmet Baki, Coiera, E., Tong, H. L., White, S. J., Quiroz, J. C., Rezazadegan, F., … Laranjo, L., "A network model of activities in primary care consultations," *Journal of the American Medical Informatics Association*, vol. *26*(10), pp. 1074–1082, 2019, doi: 10.1093/jamia/ocz046.

[54] Austin, J. L.,*How to do things with words*. Oxford university press, 1975.

[55] Grice, P., *Studies in the Way of Words*. Harvard University Press, 1989.

[56] Kalia, A., Nezhad, H. R. M., Bartolini, C., and Singh, M., "Identifying business tasks and commitments from email and chat conversations," *HP Laboratories Technical Report*, Palo Alto, CA, Rep. 4, 2013.

[57] London, A.J. *For the Common Good: Philosophical Foundations of Research Ethics*. New York, USA: Oxford University Press, 2022, 134-148.

[58] Fricker, Miranda. *Epistemic injustice: Power and the ethics of knowing*. Oxford University Press, 2007.

[59] Pettit, Philip. *Republicanism: a theory of freedom and government*. Oxford University Press, 1997.

**Alex John London** received his Ph.D. from the University of Virginia and is the Clara L. West Professor of Ethics and Philosophy at Carnegie Mellon University where he directs the Center for Ethics and Policy. His book, *For the Common Good: Philosophical Foundations of Research Ethics* was published by Oxford University Press in 2022. He is an elected fellow of the Hastings Center and served as a member of the World Health Organization (WHO) Expert Group on Ethics and Governance of AI whose report *Ethics and governance of artificial intelligence for health* was published in 2021.

**Yosef S. Razin** is a doctoral candidate in robotics at the Georgia Institute of Technology. His primary research has focused on improving our understanding and measurement of human-machine trust, especially by integrating the current approach with interdisciplinary ones from game theory and social psychology. He is a research associate in the Operational Evaluation Division at the Institute for Defense Analyses in Alexandria, VA.

**Jason Borenstein** is the Director of Graduate Research Programs within the School of Public Policy and Office of Graduate Education at the Georgia Institute of Technology. His research areas include engineering ethics, research ethics, and AI/robot ethics.

**Motahhare Eslami** earned her Ph.D. in Computer Science at the University of Illinois at Urbana-Champaign and is assistant professor at the School of Computer Science, Human-Computer Interaction Institute (HCII), and Softwared and Societal Systems Department (S3D), Carnegie Mellon University. Her research goal is to investigate the accountability challenges in algorithmic systems and to empower the users of algorithmic systems, particularly those who belong to marginalized communities or those whose decisions impact marginalized communities, make transparent, fair, and informed decisions in interaction with algorithmic systems.

**Russell Perkins** is a PhD student at the University of Massachusetts, Lowell, Francis College of Engineering where he received his B.S. and M.S. in 2018 and 2021 respectively. His research is focused on human robot trust and social robotics.

**Paul Robinette** is assistant professor in the Department of Electrical and Computer Engineering at the University of Massachusetts Lowell (UML). He has performed extensive experiments on human-robot trust in time-critical situations and has focused on field robotics: in-situ human-robot teaming experiments in the marine domain and field experiments for river navigation of autonomous surface vehicles. These projects have yielded datasets that have been released for the human-robot interaction community and the marine robotics community.