a₂ COLLECTIVE

July 31, 2023

To: Members of the PCAST Working Group on Generative AI
        Submitted via Email to: pcast@ostp.eop.gov

Re: Generative AI

The AI and Technology Collaboratories (AITCs) for Aging Research and the AITC Coordinating Center, aka **a2 Collective**,[1] appreciate the opportunity to provide input to the President's Council of Advisors on Science and Technology (PCAST) working group on generative AI. Through the following comments, we hope to contribute to PCAST's efforts to ensure a balance between encouraging innovation that translates into beneficial applications of the technology and identifying and mitigating potential harms. These timely and compelling efforts are particularly relevant to our work in soliciting proposals for promising pilot projects leveraging AI to improve care and health outcomes for older adults, including individuals with Alzheimer's disease and related dementias, and their caregivers.

We have outlined comments and recommendations that we believe will strengthen efforts to promote the beneficial deployment of generative AI. We highlight concerns regarding the added vulnerability of older adults, especially those with cognitive decline and low digital literacy, and the importance of supporting initiatives that help to socialize a culture of acceptable behavior that rewards institutional investments in training, mentorship, and accountability. We advise proceeding with caution and maintaining human oversight at both the individual and systemic levels by engaging intermediaries including caregivers and regulatory bodies to mitigate risk in older adults' interactions with generative AI, particularly in the absence of fine-tuned and validated models for this population.

In addition to the priority risks and mitigation strategies discussed in these comments, we provide in an Appendix a broader list of potential contributions generative AI can make to benefit older adults, significant risks it poses, and advisable actions to mitigate risk while maximizing the benefit of this emerging technology.

Concerns about Generative AI for Older Adults with Cognitive Decline

While we recognize potential exciting opportunities for AI to support aging and the quality of healthcare services delivered to older adults, we are particularly concerned about the use of generative AI for older adults, especially those with cognitive decline. The potential for disinformation and manipulation is high for the general public, but it becomes even more pronounced for individuals with reduced cognitive function. Among older adults, individuals with decreased

---

[1] The a2 Collective was established by the National Institute on Aging (NIA), National Institutes of Health (NIH) in 2021 to accelerate the development of AI and emerging technologies to improve the health and well-being of older adults, including individuals with Alzheimer's disease and related dementias, and their caregivers. The AITCs are centered at Johns Hopkins University (JH AITC), the University of Massachusetts, Amherst (MassAITC), and the University of Pennsylvania (PennAITech). Combined with the AITC Coordinating Center managed by Rose Li and Associates, Inc. (RLA), the **a2 Collective** supports an ecosystem that allows technologists from academic, healthcare, and private institutions to advance technology demonstration and pilot projects in partnership with multidisciplinary collaborators at these world-class research institutions. See: Abadir PM, Chellappa R, Choudhry N, et al. The promise of AI and technology to improve quality of life and care for older adults. *Nat Aging* 3, 629-631 (2023). https://doi.org/10.1038/s43587-023-0043-0

cognitive function, diminished physical and physiological well-being, or low digital literacy are already significantly more vulnerable to financial victimization compared to other groups.

One area of concern revolves around the utilization of generative AI as a solution to address loneliness, which has been acknowledged as an epidemic in a recent U.S. Surgeon General advisory.[2] Loneliness is particularly pervasive among older adults who reside alone or face limited mobility due to chronic illness. Social isolation and loneliness pose significant health risks for older adults, as they are associated with a 50 percent increase in the risk of dementia. Generative AI technologies have the potential to exacerbate the negative consequences of loneliness and risk of victimization by generating highly realistic stories that blur the lines between fact and fiction, even for well-trained individuals. The notion of "deception" as a potential feature of systems targeting individuals with cognitive limitations (for example, in the case of robotic pets) has only recently gained more attention. We are certainly lacking ethical guidelines for the appropriate use of generative AI technologies for people with mild cognitive impairment or early dementia and measures to minimize potential intended or unintended consequences of deceitful features of conversational agents. Safeguards are needed to address the challenge of informed consent with respect to the vulnerability and potential for exploitation of this population.

Considering the remarkably realistic conversations achievable with chatbots and voicebots, as well as the ability to seamlessly synthesize language, voice, and images, the development of generative AI-based highly realistic multimodal digital companions appears feasible in the near future. When used judiciously, these digital companions can offer benefits. However, there is also a notable risk that AI agents can rapidly establish trust with older adults, whose need for companionship may make them more susceptible to manipulation by malicious agents. This significantly amplifies existing challenges associated with financial fraud and misinformation.

Furthermore, even in cases where generative AI-based bots are not designed to be malicious, their advanced language capabilities and perceived trustworthiness raise questions about how these "trusted" digital companions will impact real-world interactions between older adults and their caregivers or clinicians. Will they enhance or potentially harm the relationship? The combination of generative AI and synthetic video/audio generation, enabling the mimicry of a caregiver or loved one, raises substantial concerns and may lead to long-term problems that are challenging to resolve.

In this context, it is noteworthy to consider another recent advisory from the U.S. Surgeon General regarding the effects of social media on youth mental health.[3] Just as it has taken nearly a decade to fully understand the risks associated with use of social media during adolescence, we must anticipate that it may similarly take a decade to comprehend the impact of generative AI on a generation of older adults who use smartphones and other digital tools on a daily basis and may be just starting to experience cognitive decline.

Overall, while there is significant uncertainty regarding feasible solutions given the widespread use of technology, it is evident that there is a pressing need to proceed with caution and carefully

---

[2] New Surgeon General Advisory Raises Alarm about the Devastating Impact of the Epidemic of Loneliness and Isolation in the United States
https://www.hhs.gov/about/news/2023/05/03/new-surgeon-general-advisory-raises-alarm-about-devastating-impact-epidemic-loneliness-isolation-united-states.html
[3] Surgeon General Issues New Advisory About Effects Social Media Use Has on Youth Mental Health
https://www.hhs.gov/about/news/2023/05/23/surgeon-general-issues-new-advisory-about-effects-social-media-use-has-youth-mental-health.html

consider how generative AI is employed for older adults, particularly those with cognitive decline. Further research is necessary to gain a comprehensive understanding of the appropriate guidelines and parameters governing the use of generative AI in this context.

Socializing a Culture of Acceptable Behavior that Rewards Institutional Investments in Training, Mentorship, and Accountability

To address these concerns and protect the well-being of older adults, solutions must be implemented at both the individual and systemic levels. On the individual level, caregivers play a crucial role and need to be integrated, akin to a human-in-the-loop AI decision-making model. Caregivers are generally more adept at utilizing apps created with generative AI and typically younger than the patients they assist. It is essential to ensure that apps intended for older adults with cognitive decline are designed to keep caregivers in the loop, facilitating their involvement in monitoring and decision-making. Caregivers can exercise necessary control and supervise the interactions between generative AI and older adult individuals with cognitive impairments. However, relying solely on caregivers is not a foolproof solution, as they may be first-time caregivers (such as family members) and are also susceptible to advanced manipulative tactics employed by malicious actors through generative AI. Additionally, this may further exacerbate challenges for isolated older adults who lack caregivers within their own social network.

At the caregiver and clinician levels, proactive measures can improve the use of generative AI tools for diagnosis and therapy. A well-designed and validated tool can be misused by health professionals and family caregivers if the users do not understand the scope, breadth, and derivation of the data underlying the AI model. Frail geriatric patients are as different from "the general adult population" as are pediatric patients. A generative AI system validated as appropriate for making therapy recommendations for adults aged 25-65 may make serious mistakes when unknowingly applied to frail older patients. Further, these generative AI systems may be trained on data from highly regarded studies that exclude patients with specific conditions or comorbidities. Such systems may be harmful to patients whose conditions are not represented in the data used to train the system. Just as a practicing clinician who does not keep up with new advances will provide suboptimal incorrect recommendations, a generative AI system whose knowledge base is not kept current will also proffer bad recommendations. Generative AI systems intended for clinical use should clearly indicate in what settings and with which populations they can or should not be used, when their knowledge bases were last updated, and what precautions apply.

At a systemic level, it is imperative to manage risks associated with use of generative AI, and to establish accountability measures by regulatory authorities. Any agency tasked with enforcing regulations specific to generative AI should be responsible for certifying mobile apps for older adults, ensuring that they adhere to established ethical guidelines, and monitoring efforts to synthesize language, voice, and images, particularly as multimodal digital companions are a likely use case for older adults. The integration of these modalities is crucial as full information may not be apparent from one modality alone. For effective monitoring, collaboration with companies that release foundation models is essential. Regulators should track and analyze application programming interface (API) accesses to these models, looking for patterns that might indicate malicious activity and flagging instances that may appear to violate ethical standards. Additionally, this agency should provide the infrastructure for ongoing user feedback and reporting. Channels need to be established for caregivers to provide feedback on challenging or harmful interactions and experiences, including instances of bias, misinformation, and other ethical concerns. Such a reporting mechanism provides a platform to investigate and address ongoing issues. Regulators could also engage in external auditing,

providing an objective evaluation of generative AI tools, and introduce certifications and other mechanisms to create earned trust in AI systems that they work as intended. Finally, addressing appropriate ways to provide disclosure of AI components and recognizing user expectations will be key when designing or evaluating generative AI systems specifically for older adults with cognitive impairment and for users lacking familiarity with digital tools and terminology.

Increased federal support of programs accelerating the development of ethical, trustworthy AI is also critical. Initiatives such as the a2 Collective can advance partnerships that help ensure multidisciplinary engagement, ethical standards, and inclusion of stakeholder perspectives that inform the design of generative AI systems and their application. A robust ecosystem to promote the integration of these considerations into the earliest stages of development provides a crucial corollary to the regulation of generative AI systems as they reach the market and of those already in use.

Until we acquire a complete understanding of the implications of generative AI and its potential consequences when used in the context of older adults with cognitive decline, it is crucial to enact policies that prioritize human involvement when utilizing generative AI. This ensures that human oversight is integrated to maintain ethical standards, provide transparency and appropriate support, and prevent potential harm stemming from unmonitored interactions with generative AI apps.

We would welcome any further discussions with the working group on these topics. To further access the expertise of a2 Collective principals, please contact Rose Maria Li, PhD, MBA, Director of the a2 Collective Coordinating Center, at rose.li@roseliassociates.com.

JOHNS HOPKINS
AI & TECHNOLOGY
COLLABORATORY
for AGING RESEARCH

Johns Hopkins Artificial
Intelligence and
Technology Collaboratory
for Aging Research

MassAITC

Massachusetts AI &
Technology Center for
Connected Care in Aging &
Alzheimer's Disease

PennAITech

Penn Artificial
Intelligence and
Technology
Collaboratory for
Healthy Aging

a2 COLLECTIVE
COORDINATING CENTER

Artificial Intelligence and Technology Collaboratories
for Aging Program Coordinating Center

a₂ COLLECTIVE

Appendix
Generative AI for Older Adults: Benefits, Risks, and Mitigation Strategies

Potential Beneficial Uses:

- **Personalized Data Interpretation**: Leveraging the strengths of large language models (LLMs) to distill various data sources into interpretable and actionable insights, tailored for stakeholders with varying abilities to understand and interpret information from clinicians and caregivers to patients themselves. This signifies a major advance, as LLMs can potentially personalize the conversation to different education levels, cognitive abilities, native languages, conversation styles, etc.
- **Cognitive Assistants**: Exploring the potential of cognitive assistants to aid with memory lapses, manage repetitive questions, and stabilize cognitive state at different timescales, enhancing social experiences, improving safety, and reducing caregiver stress.
- **Fraud Prevention**: Using AI as a trusted helper to provide an objective evaluation of potential fraud situations, offering added protection in an increasingly digital world.
- **Loneliness Alleviation**: Examining the capacity of highly realistic multimodal digital companions to serve as companions, addressing loneliness often experienced by older adults, especially those with cognitive decline, and potentially improving their quality of life and social engagement.

Foreseeable Risks:

- **Hyper-realistic Manipulation**: Acknowledging the risk of hyper-realistic multimodal manipulation involving video, audio, and text, which could exploit the ease with which trust can be established and potentially manipulated. Can lead to "super-scams," where mimicry of a caregiver or loved one across modalities is used to defraud vulnerable individuals.
- **Political or Behavioral Manipulation**: Increased risk of manipulating political or other actions, due to AI's capability to rapidly gain trust through high-level language fluency.

Mitigation Strategies

- **Human Oversight and Intermediaries**: Incorporating human oversight at both individual and systemic levels, engaging intermediaries such as caregivers and regulatory bodies, and advocating for policies that prioritize human involvement and oversight to ensure ethical standards, transparency, and appropriate support.
- **Stakeholder Alignment Research**: Encouraging research into alignment between AI models and the needs and preferences of caregivers, patients, clinicians, and other stakeholders, including their representation in reinforcement learning from human feedback (RLHF) and fine-tuning stages, and promoting open-source text use for fine-tuning to enhance model understanding and performance.
- **Tracking Malicious Actors**: Implementing strategies akin to cybercrime tracking to monitor API calls, detect suspicious behavior across APIs, flag potential malicious actors, and foster collaboration with foundational model-developing companies for effective monitoring.
- **Long-term Impact Study**: Supporting research aimed at understanding the long-term impact of generative AI on older adults, to enable the formulation of comprehensive guidelines and parameters for its use.
- **Regulatory Framework Development**: Calling for regulatory bodies to certify mobile apps for older adults, ensuring adherence to ethical guidelines and managing synthetic efforts, and recommending the establishment of user feedback channels and external auditing to foster trust in AI systems.
- **Ethical Guidelines Formulation**: Advocating for robust ethical guidelines, supporting initiatives that promote ethical AI development, and facilitating partnerships for multidisciplinary engagement, ethical standards, and stakeholder perspectives in the design and application of generative AI systems.