HHS Public Access

Author manuscript

J Alzheimers Dis. Author manuscript; available in PMC 2023 January 01.

Published in final edited form as:

J Alzheimers Dis. 2022 ; 88(4): 1229–1239. doi:10.3233/JAD-220441.

A Novel Explainability Approach for Technology-Driven Translational Research on Brain Aging

Adam Turnbull[1,4], **Robert Kaplan**[2], **Ehsan Adeli**[1,3], **Feng V. Lin**[1,3]

[1]Department of Psychiatry and Behavioral Sciences, Stanford University, CA, USA

[2]Clinical Excellence Research Center (CERC), Stanford University, CA, USA

[3]Wu Tsai Neurosciences Institute, Stanford University, CA, USA

[4]School of Nursing, University of Rochester Medical Center, NY, USA

Abstract

Brain aging leads to difficulties in functional independence. Mitigating these difficulties can benefit from technology that predicts, monitors, and modifies brain aging. Translational research prioritizes solutions that can be causally linked to specific pathophysiologies at the same time as demonstrating improvements in impactful real-world outcome measures. This poses a challenge for brain aging technology that needs to address the tension between mechanism-driven precision and clinical relevance. In the current opinion, by synthesizing emerging mechanistic, translational, and clinical research-related frameworks, and our own development of technology-driven brain aging research, we suggest incorporating the appreciation of four desiderata (causality, informativeness, transferability, and fairness) of explainability into early-stage research that designs and tests brain aging technology. We apply a series of work on electrocardiography-based "peripheral" neuroplasticity markers from our work as an illustration of our proposed approach. We believe this novel approach will promote the development and adoption of brain aging technology that links and addresses brain pathophysiology and functional independence in the field of translational research.

Challenges and opportunities in addressing the tension between mechanism-driven precision and clinical relevance in translational research on brain aging

Intervening brain aging to maintain or improve functional independence is among the most important goals for translational research on brain aging, and relies on understanding 1) the proposed causal biological pathways that lead to brain aging disorders in a sensitive

Corresponding author: Adam Turnbull; aturnbu2@stanford.edu.

Author contributions

Adam Turnbull: Conceptualization, Writing – original draft, Writing – Review & Editing **Robert Kaplan**: Conceptualization, Writing – Review & Editing **Ehsan Adeli:** Conceptualization, Writing – Review & Editing **Feng V. Lin:** Conceptualization, Writing – Review & Editing, Supervision, Funding acquisition

Competing interests

The authors declare that there are no competing interests.

(the disorder always acts via the pathway) and specific (the pathway does not lead to other disorders and is not present in healthy individuals) manner and 2) how these biological changes result in cognitive and functional deficits. Unfortunately, the two aspects of this goal are often in tension, highlighted by the recent controversy regarding the FDA approval of Aducanumab, a drug that showed success in reducing amyloid-beta plaques[1] (a biomarker thought to be part of the causal mechanism of Alzheimer's disease; AD) in largely white, well-educated participants without clear evidence of downstream effects on clinical outcomes of interest (e.g., cognitive decline[2]). This can be contrasted with research on physical exercise interventions, which show broad improvements in cognition across multiple brain aging disorders, often via mechanisms not specific to brain aging[3]. For example, exercise has been shown to lower chronic inflammation, which is a risk factor for AD, as well as a host of other physiological and neurological disorders, making this mechanism alone non-specific to AD. This divide reflects a broader disagreement in the human brain aging field over whether AD should be predominantly diagnosed according to biological markers thought to be involved in the causal process that leads to AD[4] or clinical outcomes, as slowing functional decline is the ultimate goal in translational AD research[5]. This tension between mechanism-driven precision and clinical relevance is partially driven by difficulties in studying brain aging in humans: early-stage research that is used to establish internal validity (i.e., establish specific and sensitive causal links to pathophysiology) uses expensive measures in controlled settings and is challenging to translate, or more specifically scale, into research in humans in the real-world that are the target of clinical tools. This has led to a disconnect between basic research studying the biological bases of brain aging and clinical research attempting to understand and improve real-world functions influenced by brain aging in humans.

Technology-driven tools used in the human brain aging field can be generally categorized into 1) brain imaging measures for understanding temporal and spatial domains of brain pathophysiology, 2) brain modulation devices for modifying brain integrity via various pathways, and 3) digital biomarkers indirectly measuring neurocognitive status (see Table 1). Compared to traditional clinical tools, these technology-driven tools can leverage advantages from autonomous, consistent data acquisition and artificial intelligence (AI) to assist in clarifying the linkage between brain and behavior to aid in allocating resources to support independence[6]. Therefore, these tools have the potential to improve 1) prediction of brain aging, characterizing *who* is at-risk earlier and with more precision, 2) monitoring of brain aging via downstream indicators and risk factors to know *when* resources are needed, and 3) modification of brain aging, establishing *how* to slow or prevent brain aging in individuals. However, to promote the development of brain aging-related technology-driven tools with both clinical relevance and mechanism-driven precision, we need to acknowledge that: 1) brain aging is highly complex, with clear links between proposed disease mechanisms and clinical outcomes remaining elusive, 2) clinical outcomes are highly heterogenous, meaning generalizability from small, homogenous samples is especially poor, and 3) methods required to establish internal validity are costly and lack scalability. The field of research aiming to develop technologies to predict, monitor, and modify brain aging is in its infancy, providing a key opportunity to improve research guidance and avoid the same mistakes and fractionation as in the broader AD literature,

particularly given that many technology developers may not appreciate the specific requirements of clinical research on brain aging.

In the current opinion paper, we emphasize a multi-dimensional understanding of explainability to guide and appraise the development of technology used in clinical research on brain aging. By leveraging our appreciation of emerging clinical research frameworks aimed at uncovering causal mechanisms (e.g., the precision medicine[7]) or intervention development (e.g., NIH stage model[8], pragmatic trial design principles[9]), we suggest that incorporating multiple desiderata of explainability into the earliest stage of technology development and testing will produce tools with both clinical relevance and mechanism-driven precision. Our goal is to promote the development and adoption of brain aging technology that links and addresses brain pathophysiology and functional independence.

The role of explainability in understanding tensions in technology-driven brain aging research

Explainability (or interpretability) refers to the extent to which the reasons why an algorithm arrived at a particular decision can be understood by humans. There is often a trade-off between predictive power and explainability: allowing AI algorithms to find the best models for predicting outcomes from input data without limiting their complexity or requiring their workings to be fully transparent to their human users often leads to improvements in their predictive power. This trade-off means that it is tempting, particularly for the developers of AI, to sacrifice explainability to maximize predictive power. However, while explainable algorithms *may* be more limited in predictive accuracy in the short term, there are long-term benefits to the generation of explainable algorithms. Notably, in clinical research, knowing which features drive predictions is critical to improving knowledge, detecting biases, facilitating social interactions, and meeting clinical standards[10]. Algorithms lacking explainability have the potential to generate predictions based on aspects of their input data that are either non-transferable or ethically problematic. For example, Watson's AI for oncology[11] showed good accuracy in controlled environments but made incorrect or even dangerous decisions when attempts were made to expand its usage into natural world settings, where the data showed complex interactions that were not present during training. Knowing which features drove its predictions could have avoided this issue. Additionally, these issues cause a lack of trust among medical professionals required to communicate decisions to patients and by patients themselves: improving explainability can improve trust by stakeholders throughout the development process, improving adoption and increasing the clinical impact of technology.

Explainability can be facilitated using inherently explainable models (e.g., transparent algorithms) or by applying post-hoc approaches to reverse-engineer solutions from more complicated algorithms (see Box 1). In translational research, these methods have to be embedded within theories of disease action to have a real impact. Understanding how different neurodegenerative diseases impact behavior in sensitive and specific ways is at the core of translational brain aging research (e.g. [4]). Thus, regulatory bodies and clinicians often prioritize explainable solutions that fit within our understanding of specific diseases

when deciding on tools to develop and adopt. In neuroscience and brain aging research, in addition to the input (e.g., neuroimages) and the output (such as the disease diagnosis), several other variables including demographics, behavioral and cognitive test scores, or genetic information are involved, and a combination of inherent and post-hoc algorithms are needed to generate appropriate and reliable explanations that link to biological pathways. For example, it is necessary to demonstrate that predictions are driven by neuroimaging signals reflecting causal biological pathways, not by biases in the demographics of the training data. Ensuring techniques are providing explainable solutions, therefore, requires additional work on the part of researchers to demonstrate not only that their algorithm is performing accurately, but that the features driving predictions can be uncovered and explained.

To provide a concrete example, AD diagnosis can be predicted with 70-90% accuracy on the basis of structural connectivity[12]. Follow-up analyses reveal which specific brain regions are driving predictions and visualize this comparatively to traditional brain imaging approaches. Doing so highlights that predictions are driven predominantly by regions known to accumulate proteins hypothesized to cause AD early on in the disease[4]. A similar analysis could be performed in the context of novel deep learning algorithms[13] for both inherent and post-hoc explanations to interpret the effect of sex, age, and other external variables on the features and predictive power of the model, to ensure these factors that lack mechanism-driven precision are not driving predictions[14]. This provides confidence both in the links between AD pathology and structural connectivity, as well as the algorithm itself. This information is helpful to researchers, clinicians, and regulatory bodies making decisions based on the development, approval, and adoption of similar technologies.

Improving explainability as a means of developing improved brain aging technology

Importantly, explainability is a multi-dimensional construct. While the emergence of techniques such as those mentioned above improves the explainability of findings identified using AI, there is no systematic way of quantifying explainability or of determining whether explanations are sufficient to specify the relationships between biological pathways of brain aging and cognitive and functional decline. For example, a cognitive neuroscience explanation of where in the brain[15] deficits of structural connectivity are most strongly related to cognitive decline may not be a sufficient explanation for a clinical evaluation of causality that is essential to know, for example, whether modifying structural connectivity would lead to improvement in cognitive function. The current tensions between mechanism-driven precision and clinical relevance in brain aging research can be understood as a mismatch in which aspects of explainability are being prioritized by stakeholders in the field[4, 5]. Lipton[16] proposes four desiderata - causality, informativeness, transferability, and fairness - critical for understanding this tension (see Table 2). Conflicts arising from mismatched explainability priorities are almost inevitable because developing tools that meet all desiderata for explainability is particularly challenging in the field of brain aging, due to trade-offs in the ability of different types of data that are prioritized at different stages of research to meet different desiderata. This means that even when tools meet one or two

explainability desiderata at one stage of research, several other desiderata are likely to be missed.

Early-stage research in controlled environments can establish causal and informative measures of brain aging pathophysiology, and link these to laboratory measures of behavior, using techniques that acquire signals directly from the brain. The most common type of causal research uses animals, leveraging the ability to directly intervene on the brain. This capacity to intervene directly on the brain allows researchers to know how the specific mechanism that they are altering causality affects 1) pathophysiology and 2) behavior; and link these two core aspects of clinical brain aging research. However, the transferability of mechanisms from animal to humans, as well as from laboratory to clinical settings are often challenging, particularly when comparing the complex interplay of real-world cognitive, social, and emotional experiences involved in clinically relevant real world behavior with laboratory measures. Both the scalability of the measures and generalizability of the findings lead to a lack of translations between studies[17]. Intervening directly on the brain in humans is more challenging and only allowable in rare cases. This makes causal links between proposed mechanisms and 1) pathophysiology and 2) behavior much more difficult to establish in humans. PET imaging can be used in early-stage research to identify potential mechanisms that cause brain aging or link brain aging with a behavioral decline in humans; however, PET is expensive and requires the injection of tracers, resulting in small, demographically homogeneous samples that are currently challenging to combine[18], limiting transferability and fairness. Genetic research can also establish potential causal pathways by identifying biological risk factors for specific brain aging disorders, but endophenotypes (e.g., measures from neuroimaging) are often required to link these informatively to clinically relevant behavior[19].

Whether the mechanisms identified in early stage research causally link pathophysiology and behavior in humans, therefore, needs to be established using progressive steps that steadily translate findings from controlled to real world settings. In clinical research, this is done most commonly using randomized controlled trials (RCTs), first in research settings, then in clinical settings, then in the real world. However, to be informative and maintain causal mechanistic links to pathophysiology, these studies require the use of neuroimaging data, including PET, EEG, and MRI, that can provide insight into whether changes seen in the behavioral outcomes are being modified via the proposed causal biological pathways. A recent perspective[20] proposed a causality continuum by which different neuroimaging studies can be evaluated, and highlighted that studies that combine multiple neuroimaging modalities (including those that involve experimental manipulation of the brain, e.g., brain stimulation) and can demonstrate “coherence” of findings across modalities provide the best evidence for causal mechanisms. These requirements to demonstrate causality place a large burden on RCTs in the field of brain aging, and the samples are often limited in size or demographic heterogeneity, as neuroimaging measures are expensive, particularly when they need to be collected over multiple timepoints as in an RCT design. This can also lead to type 2 errors due to insufficient power, making this type of research high-risk, leading many neuroimaging investigators to prioritize research in large, publicly available datasets using methods that are lower on the causality continuum (e.g., resting state fMRI). Additionally, behavioral measurements in this type of research are commonly

limited to laboratory measures (e.g. [21]) that do not generalize in clear ways to real-world behavior[22]. On the other hand, research in real-world settings can leverage autonomous, continuous behavioral data collection (e.g. from wearable devices, smartphone applications, sensors[23]) to develop transferable predictions that are more easily testable for fairness[24], but are not able to provide informative insight into the causal biological pathways that these mechanisms are involved in at the level of the brain.

To overcome this disconnect, we believe that technology developed in early-stage research needs to be appraised not only on whether it is able to establish causality and informativeness, but on whether it has the potential to demonstrate that its predictions/decisions are transferable and fair when it moves into later stage research. On the other hand, researchers developing technology aimed at real-world settings need to be made aware of the importance of causality and informativeness (in addition to transferability and fairness), and should attempt to link the improvements seen in clinical outcomes to our understanding of the biological mechanisms of brain aging. Researchers should use this general checklist to appraise the explainability of their technology, and modify their current and future research so that it meets, or has the potential to meet, all four desiderata of explainability (see Table 2). We believe that these additional checks on whether research using brain aging technology has the capacity to develop explainable tools across all four desiderata will help prioritize funding for higher impact research, allowing more researchers to utilize clinically-robust research designs. For example, research using technology within an RCT that 1) causally intervene on mechanisms with known pathways linking them to pathophysiology (informed by animal models or human PET imaging), 2) include multimodal neuroimaging markers to inform how the mechanism acts via these causal pathways, and 3) scalable measures of behavior and indicators of neural mechanisms that can be expanded to real-world research to test the transferability and fairness of these mechanisms would score highly and should be prioritized.

Incorporating multiple desiderata of explainability into early-stage technology development and testing in brain aging research

We suggest early-stage research utilize technology that can be scaled up to demonstrate transferability and fairness, enabling real-world research to be linked causally and informatively to the mechanisms of brain aging. Using the NIH stage model framework[8] as an example (see Table 3), the early stages refer to identifying components/mechanisms on which the technology (Stage 0) will be based, developing technology (Stage 1), and establishing efficacy in the laboratory (Stage 2), as opposed to later stages of establishing efficacy in clinical settings (Stage 3), showing effectiveness in the real world (Stage 4), or disseminating into the real world (Stage 5). Accordingly, several practical strategies can be applied during early-stage research: 1) research should be guided by a clear clinical premise, aimed at predicting, monitoring, or modifying a specific and sensitive brain aging pathway informed by stage 0 research and 2) attempts should be made to incorporate real-world measures into early-stage research, to identify coherence across data modalities that can establish internal validity (i.e., neuroimaging) and those that can be used to establish external validity (i.e., wearable devices). Using the four desiderata of explainability can

help researchers to judge whether their research is fostering, or will be able to foster, explainability that can be embedded into theories of disease action: can features driving predictions be causally and informatively linked to both pathophysiology and behavior that is real-world relevant, and are they transferable and fair. In early stages, it is appropriate to prioritize causality and informativeness, however, researchers need to consider whether their technology is feasible in terms of ascertaining transferability and fairness at later stages, and incorporating scalable data modalities into early-stage research is the best way of ensuring this. Features based on these scalable modalities can then be moved through to late-stage research, knowing that they have demonstrable links to pathophysiology. Similarly, researchers who have identified transferable and fair behavioral markers should work on building these back into early-stage research to determine causal and informative links to pathology.

Model predictions should be clearly linked to the clinical features that drive them, facilitated by the use of inherently explainable algorithms or post-hoc approaches to reverse engineering explanations[10]. Research designs should prioritize maximizing explainability desiderata: e.g., using RCTs and referring to the causality continuum in neuroimaging studies[20] to maximize causality, basing hypotheses on, and attempting to interpret, potential biological mechanisms to improve informativeness, using training, test, validation methods in large, diverse samples using real-world measures to improve transferability, collecting demographically diverse samples, and specifically accounting for potential socioeconomic or ethnic biases to ensure fairness. Meeting all desiderata of explainability would benefit from collaborative research between experts from multiple fields within brain aging research. Animal models, for example, are essential for stage 0 research to identify novel candidate mechanisms on which technology can be built. Clinical scientists and psychiatrists are best able to design and implement RCTs in order to establish the causal basis for these technologies in difficult-to-study patient groups, and neuroscientists will be needed to maximize the informativeness of the mechanisms that the technology leverages. This will require a systems neuroscience approach, building teams with expertise at different levels of understanding the same biological pathways to establish coherence between the different levels of measurement that are required for the four desiderata of explainability. Additionally, ethicists and social scientists will be needed to ensure studies testing transferability and fairness take into account all potential sources of bias. To facilitate this, researchers should also strive to meet FAIR guiding principles for scientific data management and stewardship[25], allowing researchers to foster explainability collaboratively, adding causal and informative links to real-world biomarkers, and developing scalable markers to help establish transferability and fairness of behavioral mechanisms of known indicators of pathology.

Example research: technology-measured adaptation capacity as a target for cognitive training

To illustrate a practical example of the incorporation of explainability into the early stages of designing and testing a technology-driven brain aging tool, we outline how research into a specific brain aging process – adaptation capacity – can be leveraged to develop technology

to modify brain aging with comprehensive explainability using research from our lab (see Table 2). We highlight how we believe our approach helps to improve the clinical relevance of this particular mechanism, while maintaining mechanism-driven precision.

Early stage research (e.g., mechanistic studies from stage 0) has established that the connectivity of specific regions of the brain, particularly those in the cingulate cortex, is maintained in older adults with superior memory[26], and that this connectivity may confer protection against AD pathology[27]. Further stage 0 research has shown that autonomic nervous system (ANS) function relates to cingulate cortex function[28], and ANS responses to cognitive training are important for determining whether individuals show improvement following the intervention[29]. This establishes a biological pathway via which ANS responses to cognitive training may intervene on brain networks known to be essential for AD, allowing for technology developed to leverage this pathway to have mechanism-driven precision, if future research can establish how this pathway modifies AD pathophysiology. Importantly, measurement of the ANS using ECG is highly scalable and can be done with wearable devices that give this potential intervention high clinical relevance if it can be leveraged effectively to improve cognitive training outcomes.

This research, therefore, inspired intervention development and pilot testing research (stage 1) to test whether cognitive training can improve ANS function via this shared mechanism involving the cingulate cortex[30], and whether improving ANS function can further strengthen the effect of cognitive training in cognitive aging[31]. Separately, in a stage 2 trial, we also revealed that selected patterns of ANS function can predict neuroplasticity in the cingulate cortex following training[32]. This study used an inherently explainable machine-learning approach called shapelet analysis with a limited number of theory-derived features to provide clear explanations for predictions that could be incorporated into findings from stage 0 research to develop more comprehensive explainability. This knowledge can feedback to develop new tools and cognitive training paradigms: e.g., early-stage research directly monitoring and modifying adaptation capacity in an individualized cognitive training[33]. Additional research is required to understand the precise mechanism by which ANS function relates to AD pathophysiology and how interventions that act on adaptation capacity might causally alter AD pathophysiology, potentially using an RCT design alongside multimodal neuroimaging. The fact that ANS measures are scalable allows for later stage research, potentially leveraging app-based cognitive training approaches alongside wearable ANS measures, to determine the transferability and fairness of these biomarkers. Attention will need to be paid to the fact that some wearable ANS measures are known to be less accurate in individuals with darker skin tones[24], to ensure this biomarker is effective in ethnic minority populations and doesn’t widen healthcare outcome disparities. If successful, research based on theories surrounding adaptation capacity could result in scalable technology to modify brain aging that meets all desiderata for explainability: cognitive training that improves cognition based on mechanisms with 1) causal, informative links to the resistance to AD pathology and 2) transferability and fairness that ensure this technology is equally successful in the real world in all individuals. If research is carried out according to this framework, this technology would have both mechanism-driven precision: an understanding of the biological pathways via which is causally alters AD

pathophysiology, as well as clinical relevance: fair and transferable effects that improve cognitive function in the real world in older adults at-risk for AD.

Conclusion

We here urge the field of translational research on brain aging to recognize the tension between mechanism-driven precision and clinical relevance in existing research trying to predict, monitor, and modify brain aging and associated functional outcomes. Technology-driven tools provide the potential to help link these goals that have become disconnected in the broader AD literature, but only if individuals developing technology are aware of, and take active steps to mitigate, this tension. Utilizing the four desiderata of explainability to judge the potential of clinical technology *early* on in the design process can help to identify potential explainability issues later on, and may therefore help promote the dissemination of cost-effective technology to promoting successful brain aging.

Acknowledgement:

The authors have no conflicts of interest. The paper development was supported by NIH/NIA R21/R33 AG073356 and NIH/NINR R01 NR015452B to Feng V. Lin.

References

[1]. Sevigny J, Chiao P, Bussière T, Weinreb PH, Williams L, Maier M, Dunstan R, Salloway S, Chen T, Ling Y (2016) The antibody aducanumab reduces Aβ plaques in Alzheimer's disease. Nature 537, 50–56. [PubMed: 27582220]

[2]. Selkoe DJ (2019) Alzheimer disease and aducanumab: adjusting our approach. Nature Reviews Neurology 15, 365–366.

[3]. Kennedy G, Hardman RJ, Macpherson H, Scholey AB, Pipingas A (2017) How does exercise reduce the rate of age-associated cognitive decline? A review of potential mechanisms. Journal of Alzheimer's Disease 55, 1–18.

[4]. Jack CR Jr, Bennett DA, Blennow K, Carrillo MC, Dunn B, Haeberlein SB, Holtzman DM, Jagust W, Jessen F, Karlawish J (2018) NIA-AA research framework: toward a biological definition of Alzheimer's disease. Alzheimer's & Dementia 14, 535–562.

[5]. Garrett MD (2018) A Critique of the 2018 National Institute on Aging's. Research Framework: Toward a biological definition of Alzheimer's disease. Curr Neurobiol 9, 49–58.

[6]. Astell AJ, Bouranis N, Hoey J, Lindauer A, Mihailidis A, Nugent C, Robillard JM (2019) Technology and dementia: The future is now. Dementia and geriatric cognitive disorders 47, 131–139. [PubMed: 31247624]

[7]. Collins FS, Varmus H (2015) A new initiative on precision medicine. N Engl J Med 372, 793–795. [PubMed: 25635347]

[8]. Onken LS (2019) History and evolution of the NIH stage model. Evidence–based practice in action: bridging clinical science and intervention, 28–42.

[9]. Steinman MA, Boyd CM, Schmader KE (2021) Expanding Evidence for Clinical Care of Older Adults: Beyond Clinical Trial Traditions and Finding New Approaches. JAMA 326, 475–476. [PubMed: 34292309]

[10]. Molnar C (2019) Interpretable machine learning. A Guide for Making Black Box Models Explainable, https://christophm.github.io/interpretable-ml-book/.

[11]. Strickland E (2019) IBM Watson, heal thyself: How IBM overpromised and underdelivered on AI health care. IEEE Spectrum 56, 24–31.

[12]. Essemlali A, St-Onge E, Descoteaux M, Jodoin P-M (2020) in Medical Imaging with Deep Learning PMLR, pp. 217–229.

[13]. Adeli E, Zhao Q, Zahr NM, Goldstone A, Pfefferbaum A, Sullivan EV, Pohl KM (2020) Deep learning identifies morphological determinants of sex differences in the pre-adolescent brain. NeuroImage 223, 117293. [PubMed: 32841716]

[14]. Zhao Q, Adeli E, Pohl KM (2020) Training confounder-free deep learning models for medical applications. Nature communications 11, 1–9.

[15]. van Bree S (2022) A critical perspective on neural mechanisms in cognitive neuroscience: towards unification.

[16]. Lipton ZC (2018) The Mythos of Model Interpretability: In machine learning, the concept of interpretability is both important and slippery. Queue 16, 31–57.

[17]. Götz J, Bodea L-G, Goedert M (2018) Rodent models for Alzheimer disease. Nature Reviews Neuroscience 19, 583–598. [PubMed: 30194347]

[18]. Schwarz CG (2021) Uses of Human MR and PET Imaging in Research of Neurodegenerative Brain Diseases. Neurotherapeutics, 1–12. [PubMed: 33821447]

[19]. Walters JTR, Owen MJ (2007) Endophenotypes in psychiatric genetics. Molecular psychiatry 12, 886–890. [PubMed: 17895920]

[20]. Siddiqi SH, Kording KP, Parvizi J, Fox MD (2022) Causal mapping of human brain function. Nature Reviews Neuroscience, 1–15. [PubMed: 34754097]

[21]. Mueller SG, Weiner MW, Thal LJ, Petersen RC, Jack CR, Jagust W, Trojanowski JQ, Toga AW, Beckett L (2005) Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's Disease Neuroimaging Initiative (ADNI). Alzheimer's & Dementia 1, 55–66.

[22]. Nastase SA, Goldstein A, Hasson U (2020) Keep it real: rethinking the primacy of experimental control in cognitive neuroscience. NeuroImage 222, 117254. [PubMed: 32800992]

[23]. Kim K-i, Gollamudi SS, Steinhubl S (2017) Digital technology to enable aging in place. Experimental gerontology 88, 25–31. [PubMed: 28025126]

[24]. Colvonen PJ (2021) Response To: Investigating sources of inaccuracy in wearable optical heart rate sensors. NPJ Digital Medicine 4, 1–2. [PubMed: 33398041]

[25]. Stall S, Yarmey L, Cutcher-Gershenfeld J, Hanson B, Lehnert K, Nosek B, Parsons M, Robinson E, Wyborn L (2019) Nature Publishing Group.

[26]. Lin F, Ren P, Mapstone M, Meyers SP, Porsteinsson A, Baran TM, Initiative AsDN (2017) The cingulate cortex of older adults with excellent memory capacity. Cortex 86, 83–92. [PubMed: 27930899]

[27]. Lin F, Ren P, Lo RY, Chapman BP, Jacobs A, Baran TM, Porsteinsson AP, Foxe JJ (2017) Insula and inferior frontal gyrus' activities protect memory performance against Alzheimer's disease pathology in old age. Journal of Alzheimer's Disease 55, 669–678.

[28]. Lin F, Ren P, Wang X, Anthony M, Tadin D, Heffner KL (2017) Cortical thickness is associated with altered autonomic function in cognitively impaired and non-impaired older adults. J Physiol 595, 6969–6978. [PubMed: 28952161]

[29]. Lin F, L Heffner K, Ren P, Tadin D (2017) A role of the parasympathetic nervous system in cognitive training. Current Alzheimer Research 14, 784–789. [PubMed: 28164771]

[30]. Lin FV, Tao Y, Chen Q, Anthony M, Zhang Z, Tadin D, Heffner KL (2020) Processing speed and attention training modifies autonomic flexibility: A mechanistic intervention study. Neuroimage 213, 116730. [PubMed: 32165263]

[31]. Lin FV, Heffner K, Gevirtz R, Zhang Z, Tadin D, Porsteinsson A (2021) Targeting autonomic flexibility to enhance cognitive training outcomes in older adults with mild cognitive impairment: study protocol for a randomized controlled trial. Trials 22, 560. [PubMed: 34425878]

[32]. Chen Q, Yang H, Rooks B, Anthony M, Zhang Z, Tadin D, Heffner KL, Lin FV (2020) Autonomic flexibility reflects learning and associated neuroplasticity in old age. Human brain mapping 41, 3608–3619. [PubMed: 32510759]

[33]. Lin F, Zhang Z, Tapparello C (2021-2026) NIH/NIA R21/R33 AG073356

[34]. Okamura N, Harada R, Furumoto S, Arai H, Yanai K, Kudo Y (2014) Tau PET imaging in Alzheimer's disease. Current neurology and neuroscience reports 14, 500. [PubMed: 25239654]

[35]. Fernández A, Hornero R, Mayo A, Poza J, Gil-Gregorio P, Ortiz T (2006) MEG spectral profile in Alzheimer's disease and mild cognitive impairment. Clinical Neurophysiology 117, 306–314. [PubMed: 16386951]

[36]. Li HJ, Hou XH, Liu HH, Yue CL, He Y, Zuo XN (2015) Toward systems neuroscience in mild cognitive impairment and Alzheimer's disease: A meta-analysis of 75 fMRI studies. Human Brain Mapping 36, 1217–1232. [PubMed: 25411150]

[37]. Leandrou S, Petroudi S, Kyriacou PA, Reyes-Aldasoro CC, Pattichis CS (2018) Quantitative MRI brain studies in mild cognitive impairment and Alzheimer's disease: a methodological review. IEEE reviews in biomedical engineering 11, 97–111. [PubMed: 29994606]

[38]. Clerx L, Visser PJ, Verhey F, Aalten P (2012) New MRI markers for Alzheimer's disease: a meta-analysis of diffusion tensor imaging and a comparison with medial temporal lobe measurements. Journal of Alzheimer's Disease 29, 405–429.

[39]. Dennis EL, Thompson PM (2014) Functional brain connectivity using fMRI in aging and Alzheimer's disease. Neuropsychology review 24, 49–62. [PubMed: 24562737]

[40]. Horvath A, Szucs A, Csukly G, Sakovics A, Stefanics G, Kamondi A (2018) EEG and ERP biomarkers of Alzheimer's disease: a critical review. Frontiers in bioscience (Landmark edition) 23, 183–220. [PubMed: 28930543]

[41]. Rahman MA, Siddik AB, Ghosh TK, Khanam F, Ahmad M (2020) A narrative review on clinical applications of fNIRS. Journal of Digital Imaging, 1–18. [PubMed: 32076923]

[42]. Nardone R, Tezzon F, Höller Y, Golaszewski S, Trinka E, Brigo F (2014) Transcranial magnetic stimulation (TMS)/repetitive TMS in mild cognitive impairment and Alzheimer's disease. Acta Neurologica Scandinavica 129, 351–366. [PubMed: 24506061]

[43]. Hescham S, Lim LW, Jahanshahi A, Blokland A, Temel Y (2013) Deep brain stimulation in dementia-related disorders. Neuroscience & Biobehavioral Reviews 37, 2666–2675. [PubMed: 24060532]

[44]. Birba A, Hesse E, Sedeño L, Mikulan EP, García MdC, Ávalos J, Adolfi F, Legaz A, Bekinschtein TA, Zimerman M (2017) Enhanced working memory binding by direct electrical stimulation of the parietal cortex. Frontiers in Aging Neuroscience 9, 178. [PubMed: 28642698]

[45]. Freitas C, Mondragón-Llorca H, Pascual-Leone A (2011) Noninvasive brain stimulation in Alzheimer's disease: systematic review and perspectives for the future. Experimental gerontology 46, 611–627. [PubMed: 21511025]

[46]. Luijmes RE, Pouwels S, Boonman J (2016) The effectiveness of neurofeedback on cognitive functioning in patients with Alzheimer's disease: Preliminary results. Neurophysiologie Clinique/Clinical Neurophysiology 46, 179–187. [PubMed: 27374996]

[47]. Trambaiolli LR, Cassani R, Mehler D, Falk TH (2021) Neurofeedback and the aging brain: a systematic review of training protocols for dementia and mild cognitive impairment. Frontiers in aging neuroscience 13, 270.

[48]. Tang Y, Lin X, Lin X-j, Zheng W, Zheng Z-k, Lin Z-m, Chen J-h (2017) Therapeutic efficacy of neuromuscular electrical stimulation and electromyographic biofeedback on Alzheimer's disease patients with dysphagia. Medicine 96.

[49]. Templeton JM, Poellabauer C, Schneider S (2020) Enhancement of neurocognitive assessments using smartphone capabilities: systematic review. JMIR mHealth and uHealth 8, e15517. [PubMed: 32442150]

[50]. Kourtis LC, Regele OB, Wright JM, Jones GB (2019) Digital biomarkers for Alzheimer's disease: the mobile/wearable devices opportunity. NPJ digital medicine 2, 1–9. [PubMed: 31304351]

[51]. Gillani N, Arslan T (2021) Intelligent Sensing Technologies for the Diagnosis, Monitoring and Therapy of Alzheimer's Disease: A Systematic Review. Sensors 21, 4249. [PubMed: 34205793]

[52]. Lu M, Zhao Q, Poston KL, Sullivan EV, Pfefferbaum A, Shahid M, Katz M, Kouhsari LM, Schulman K, Milstein A (2021) Quantifying Parkinson's disease motor severity under uncertainty using MDS-UPDRS videos. Medical Image Analysis 73, 102179. [PubMed: 34340101]

[53]. Pistollato F, Ohayon EL, Lam A, Langley GR, Novak TJ, Pamies D, Perry G, Trushina E, Williams RS, Roher AE (2016) Alzheimer disease research in the 21st century: past and current failures, new perspectives and funding priorities. Oncotarget 7, 38999. [PubMed: 27229915]

[54]. Anderson NB, Bulatao RA, Cohen B, on Race P, Council NR (2004) Ethnic differences in dementia and Alzheimer's disease In Critical perspectives on racial and ethnic differences in health in late life National Academies Press (US).

[55]. Van De Vorst IE, Koek HL, Stein CE, Bots ML, Vaartjes I (2016) Socioeconomic disparities and mortality after a diagnosis of dementia: results from a nationwide registry linkage study. American journal of epidemiology 184, 219–226. [PubMed: 27380760]

Box 1.

Getting explainable solutions from machine learning

Inhrently explainable algorithms

Inherently explainable algorithms have model internals (e.g. learned weights) that are human interpretable and can be easily understood. Inherently explainable algorithms include, but are not limited to, linear regression, logistic regression, generalized linear or additive models (GLMs and GAMs), and decision trees or rules, gradient boosting (see [10] for full review). Developing machine learning approaches that are inherently explainable may also involve limiting model complexity as even simple approaches (e.g., linear regression) become unexplainable if the number of features increases drastically. Thinking clearly about how to visualize solutions can improve explainability.

Post-hoc methods

Techniques also exist to reverse-engineer trained models, even those generated by "black-box" algorithms. This allows researchers to take advantage of more powerful (complex) algorithms. For example, varying input features and measuring outputs can produce weights that can be explained and presented in a similar way to those from inherently explainable algorithms. These approaches include partial dependence plots, individual conditional expectation, accumulated local effects plots, Shaply values, as well as others (see [10] for full review). Recent deep learning algorithms also use saliency map generation methods as a means of explainaing neural network decisions (e.g., [13]).

Table 1.

Literature on technology-driven human brain aging research

Category	Technology used in human brain aging research
Brain imaging measures for temporal and spatial domains of brain pathophysiology and function	• PET for the spatial localization of proteins associated with brain aging[34] • MEG for understanding the fast temporal dynamics (via electrical signals) of brain function with some spatial specificity[35] • Multi-modality MRI for understanding slow functional dynamics (functional MRI[36]; via hemodynamic signals) in brain activity with high spatial specificity, and for understanding the structural (structural MRI[37] and DTI[38]) and functional organization of the brain (resting state functional MRI[39]) • EEG for understanding the fast temporal dynamics (via electrical signals) of brain activity from the surface of the scalp with limited spatial specificity[40] • fNIRs for understanding the slow temporal dynamics (via hemodynamic signals) of brain activity from regions near the surface with increased portability and scalability than fMRI[41]
Brain modulation devices for modifying brain integrity	• TMS for producing electrical stimulation on brain regions under the surface of the scalp using magnetic fields[42], or DBS/intracranial stimulation to stimulate deeper regions more directly[43, 44] • tDCS for producing electrical stimulation between two brain regions under the surface of the scalp using electrical currents[45] • Real-time neuroimaging (or neurofeedback) during which participants are provided with their own brain signals and asked to increase or decrease these using their own cognition[46, 47] • Biofeedback during which participants are provided with their own peripheral signals reflecting brain activity and asked to increase or decrease these using their own cognition or behavior[48]
Digital biomarkers indirectly measuring neurocognitive statuses	• Smartphone applications that can be used to display cognitive or behavioral tests[49], (e.g., BGC Science from UCR to measure episodic/semantic memory and executive functions) • Wearable devices for measuring participant behavior: movement, heart rate, blood pressure, etc.[50], (e.g., smartwatches or smart rings) • Sensors that can detect participant behavior from an external source (e.g., in-home cameras)[51] along with computer vision methods that can automatically find body markers highly correlated with a neurological disease[52]

NOTE: DTI = diffusion tensor imaging; EEG = Electroencephalography; fNIRs = Functional near-infrared spectroscopy; MEG = Magnetoencephalography; MRI = magnetic resonance imaging; PET = positron emotion tomography; tDCS = transcranial direct current stimulation; TMS = transcranial magnetic stimulation;

Table 2.

Overview of explainability desiderata in brain aging research

Desiderate	Overview	Definition within brain aging research	Evaluation criteria for guiding the development of technology for brain aging research	Examples from technology-measured adaptation capacity for preventing brain aging and functional decline
Causality	Determine if pathophysiology causes outcome?	Determinizing that specific pathophysiologies cause specific brain aging diseases is essential to clinical brain aging research. Causal links to pathophysiology require expensive techniques and controlled settings to establish. Some researchers believe that an over-emphasis on ensuring tools affect the causal biological basis for AD may be causing "surrogate markers" (e.g., amyloid-beta plaques) to become the focus of research, rather than a means to improving clinical outcomes[2, 5].	Does/could this technology show that it can causally modulate the biological mechanisms of brain aging for the specific disease of interest?	A study identified baseline adaptation capacity (reflected in a heart rate shapelet in response to cognitive demands) was associated with later improvements in cognitive function at 7 weeks following cognitive training[32]. To establish that adaptation capacity is causally related to cognitive training improvements, a follow-up randomized controlled trial has been proposed that targets adaptation capacity in the experimental group [31]. The comparison of this group with an active control will allow a clearer understanding of the causal role of adaptation capacity in cognitive training improvements. Additional research is also required to understand the causal relationship between this biomarker and pathophysiology that may require PET imaging or the extraction of blood-based biomarkers.
Informativeness	Does technology inform future research?	Maximizing informativeness improves knowledge of the causes and outcomes of brain aging, facilitating the improvement and development of future tools. Different stages of research produce different levels of informativeness, but biological mechanisms that can be linked to pathophysiology are prioritized by current funding schemes, particularly for early-stage research[53].	Does/could this technology advance our understanding of the specific brain aging mechanisms we are researching?	Baseline autonomic flexibility was shown to be related to training-induced flexibility in the anterior cingulate cortex, a region known to be important for cognitive function in AD[32].
Transferability	Do results generalize?	Transferability ensures that decisions/predictions generalize to non-analyzed participants and contexts. It is necessary to limit mistakes and biases caused by increasing complexity as research moves into real-world settings[11]. Large, diverse samples with externally valid measures are better able to demonstrate transferability, and these are currently not common in brain aging research, particularly at early stages. Collecting these samples requires the use of scalable measures, and these need to be prioritized if transferability is to be established.	Are the predictions/decisions made by this technology generalizable beyond the setting and sample currently being studied? Is it feasible for future studies to demonstrate that predictions/decisions generalize (i.e., are metrics scalable, or do they have reasonable, scalable proxies)?	Both the cognitive training and heart rate data from the randomized controlled trial can be scaled up via smartphones and wearable devices. The ability of these devices to reproduce the same biomarker needs to be established in future research. It is important to note that these studies do not establish the transferability of these mechanisms, and suffer from similar issues to the broader field, particularly low sample size and heterogeneity. However, the inclusion of scalable measurements techniques allows for a greater potential for future research to establish whether these mechanisms are transferable than if only neuroimaging was used.
Fairness	Does research information contribute to widening healthcare disparities?	Ensuring fairness is necessary to ensure medical decisions are not biased and likely to reinforce healthcare disparities, taking into account that dementia rates are elevated amongst less educated, poorer, and ethnic minority populations[54, 55]. Large, diverse samples are crucial to establishing fairness, and these have been lacking in brain aging research to date.	Do the predictions/decisions made by this technology have the potential to widen healthcare disparities? What steps need to be taken to ensure these findings are equally beneficial to all individuals, particularly those from ethnic minorities and socioeconomically disadvantaged backgrounds, or heterogeneous healthcare systems? Are these steps feasible?	The ability of the heart rate shapelet biomarker to be tested in the real-world using smartphones and wearable devices allows for large, diverse samples to establish whether predictions hold in ethnic minorities and individuals from socioeconomically disadvantaged individuals, and in diverse healthcare systems.

Table 3.

Proposed research stages (Note: there is significant feed-forward and feed-back between stages, and research does not neccesarily have to follow the stages precisely)

Stage (NIH stage model)	Aim of research	Research designs in brain aging research	Modifications to promote explainability
Early stage			
Stage 0: pre-intervention	Identify potential mechanisms	Animal models, human PET imaging, neuroimaging with biomarkers	Potential mechanisms should show robust causal relationships to pathways linking brain aging pathophysiology and behavior, ideally with evidence that they relate to both these core aspects
Stage 1: intervention creation and testing	Create intervention based on identified mechanisms and perform feasibility and pilot testing	Pilot/preliminary RCTs in research settings	Research designs should include neuroimaging measures to inform mechanistic precision as well as scalable measures of behavior and downstream indicators of mechanistic change to allow for scalability
Stage 2: pure efficacy	Experimental testing of interventions in research settings with research-based providers	Full RCTs in research settings	In addition to maintaining modifications from stage 1 research, attempts should be made to recruit large, heterogenous samples that can demonstrate some degree of transferability and fairness of mechanisms. Alternatively, research should include trials specifically involving underrepresented populations to test generalizability of mechanisms
Later stage			
Stage 3: real-world efficacy	Experimental testing of interventions in community setting with community providers	RCTs in community settings, pragmatic trials	Transferability and fairness of mechanisms should be prioritized, scalable measures with links to pathophysiology from early stage research can allow for mechanistic precision without the need for expensive neuroimaging that limits scalability
Stage 4: effectiveness	Examines empirically supported interventions in community settings with community providers	Cohort studies, cross-sectional research	Use of scalable markers from previous stages allows for transfer of mechanistic precision with a known causal basis even in cross-sectional research, allowing for maximum sample sizes to determine transferability and fairness
Stage 5: implementation and dissemination	Examines strategies of implementation and dissemination of empirically supported inverventions	Cohort studies, cross-sectional research with a focus on engagement	Attempts should be made to modify technology to maximize dissemination and impact. Links may need to be established between novel devices with improved scalability and those used in early stage research to ensure no loss of mechanistic precision while improving clinical impact