nature aging

Perspective

https://doi.org/10.1038/s43587-026-01186-z

Agentic AI for scaling diagnosis and care in neurodegenerative disease

Received: 8 July 2025

Accepted: 1 July 2026

Published online: 30 July 2026

Check for updates

Andrew G. Breithaupt[1] ✉, Michael Weiner[2], Alice Tang[3,4], Katherine L. Possin[5,6], Marina Sirota[4], James Lah[1], Allan I. Levey[1], Pascal Van Hentenryck[7,8], Reza Zandehshahvar[7,13], Marilu Luisa Gorno-Tempini[5,6], Joseph Giorgio[9], Jingshen Wang[10], Andreas M. Rauschecker[2,11], Howard J. Rosen[5,6], Rachel L. Nosheny[2,12], Bruce L. Miller[5,6] & Pedro Pinheiro-Chagas[5] ✉

US healthcare systems are struggling to meet the growing demand for neurological care, particularly in Alzheimer's disease and related dementias. Generative artificial intelligence (AI) built on large language models now enables agentic AI systems that can streamline clinical workflows, integrate multimodal data and learn from practicing specialists. We envision an agentic AI system that scales specialist-level care to nonspecialist clinical settings through a continuously learning healthcare system. We describe this destination and outline a phased roadmap for responsible design and integration into care of Alzheimer's disease and related dementias: (1) high-quality standardized data collection across modalities; (2) decision support; (3) clinical integration enhancing workflows; (4) rigorous validation and monitoring protocols; (5) continuous learning through clinical feedback; and (6) robust ethics and risk management frameworks. This human-centered approach optimizes clinicians' capabilities in comprehensive data collection, interpretation of complex clinical information and timely application of relevant medical knowledge while prioritizing patient safety, healthcare equity and transparency.

As the US population ages, the number of people with Alzheimer's disease and related dementias (ADRD) is rapidly increasing[1,2], with US cases projected to double by 2060 to 1 million annually[3]. Early detection and precise diagnosis are crucial for proper access to quality care, eligibility for new and emerging disease-modifying therapies and enabling research to advance the field. Unfortunately, more than 50% of dementia diagnoses are delayed until moderate or advanced stages in primary care[4,5] with greater delays among racial and ethnic minority individuals[6].

There are many challenges with diagnosing neurodegenerative diseases. Patients usually go to their primary care providers first, but most primary care providers cite lack of time and confidence with ADRD diagnosis[7,8]. Referrals to dementia specialists face substantial barriers as the demand greatly exceeds the available specialist supply with the gap continuing to widen and wait times continuing to grow[9–12]. Beyond access issues, diagnosis is challenging, requiring time-consuming data collection and nuanced interpretation. These challenges are amplified by exponentially growing data complexity, and an overwhelming volume of medical literature and research opportunities that exceed any individual clinician's capacity to master.

After diagnosis, challenges with management are growing as new therapeutics with intensive requirements for monitoring are emerging[13,14]. Many primary care providers and general neurologists struggle to keep pace with the rapid advances in the field, including maintaining awareness of new treatments, diagnostics and research studies that could potentially benefit their patients and the ADRD field. Furthermore, new care models such as those proposed by Centers for Medicare and Medicaid Services (CMS)[15] demand additional resources that our current workforce is ill-equipped to support.

A full list of affiliations appears at the end of the paper. ✉e-mail: andrew.g.breithaupt@emory.edu; pedro.pinheirochagas@ucsf.edu

Emerging solutions from advances in AI

Simultaneously, AI is rapidly advancing and has entered a new era through generative AI built on large language models (LLMs). A growing number of studies have demonstrated that LLM-based systems encode clinical knowledge[16], can achieve expert-level medical question answering[17], integrate multimodal patient data[18], are capable of empathetic history-taking through conversational agents[19] and enhance clinical decision-making[20]. Critically, clinicians are voluntarily and rapidly adopting these tools: a 2026 American Medical Association survey found that 81% of physicians now use AI tools in practice, more than double the rate in 2023 (ref. 21). Unlike prior clinical decision support (CDS) systems that struggled with adoption, LLM-based tools are being actively sought out by clinicians, fundamentally changing the feasibility of AI-integrated healthcare. However, single LLMs face important limitations in clinical deployment, including hallucinations, inability to learn continuously in the complicated healthcare landscape and limited capacity to integrate multimodal data (Box 1).

LLMs have enabled agentic AI systems[22,23] (Box 2). These systems can search medical databases, analyze structured and unstructured patient data, and connect to electronic health records (EHRs). Although much research has focused on AI's potential to identify patterns beyond human perception (predictive AI), clinical implementation remains challenging because clinicians cannot readily verify insights they cannot detect, nor can these systems naturally complement clinical workflows. However, agentic AI systems offer more immediate, practical applications by enhancing clinicians' capabilities in three critical areas where healthcare providers increasingly struggle: comprehensive data collection, interpretation of increasingly complex multimodal information, and timely, context-appropriate application of relevant medical knowledge. Importantly, because clinicians are already engaging with LLM-based tools in routine practice, agentic systems built on this foundation can leverage existing adoption patterns rather than requiring new behavioral change.

Currently, we can deploy domain-specific AI assistants for individual functions, each optimized for its particular role in data collection, interpretation or knowledge retrieval (Fig. 1a); however, the future vision involves an integrated agentic AI system capable of seamlessly coordinating these functions within a continuously learning healthcare framework (Fig. 1b). Such systems would leverage every patient's encounter while incorporating the latest clinical evidence. Although agentic AI systems of this scope do not yet exist in clinical practice, the rapid convergence of clinician adoption, LLM capabilities in leveraging unstructured clinical data and advances in multimodal integration create a pathway towards this goal.

Still, such a learning system must be intentionally molded to benefit diverse patient populations, support both patients and providers without imposing new burdens and be safely implemented in any clinic setting. Recent concerns around generative and agentic AI have emphasized the need for strict boundaries around autonomy and for maintaining robust human oversight in high-stakes medical contexts[24]. This is particularly important in ADRD, where vulnerable populations, shifting decision-making capacity and high-stakes diagnostic uncertainty could magnify risk. Therefore, we advocate for constrained agentic AI systems in which clinicians remain in the loop, supported by rigorous evaluation and continuous monitoring using clinically meaningful outcomes.

In this Perspective, we envision an agentic AI-enabled learning healthcare system with the potential to scale specialist-level clinical care in ADRD in the USA (Fig. 1b), and we outline a roadmap to this destination (Table 1) across six interconnected phases, drawing on AI implementation principles adapted to the unique attributes of agentic systems. In the sections that follow, we first discuss the foundation of high-quality standardized multimodal data collection, which enables downstream AI capabilities. We then describe how an agentic AI decision support system can synthesize this data to produce evidence-based recommendations a clinician can interpret and verify. We next address clinical integration and workflow design, followed by the validation and monitoring necessary to ensure safety and maintain efficacy. Finally, we examine how this system enables continuous learning, and end with an examination of ethics and risk management considerations that guide responsible deployment. Although our roadmap presents clinical integration as a distinct phase, implementation success requires designing all components with clinical workflows in mind from the outset. Therefore, integration considerations are intentionally woven throughout this paper. We hope this framework will prove useful to anyone in the field of neurology or other specialties, encouraging exploration into how agentic AI may aid their patients and fellow providers in ways previously not possible.

BOX 1

Limitations of single LLMs in clinical practice

Single LLMs have demonstrated impressive performance on standardized medical examinations and diagnostic benchmarks[16,17], and clinicians are adopting LLM-powered tools at a historically unprecedented rate[21]. Yet several structural limitations constrain their clinical utility in practice.

LLMs exhibit what has been termed 'jagged intelligence': strong performance on some complex reasoning coexists with unpredictable failures on other simpler tasks, making reliability difficult to anticipate in novel clinical situations[107]. Hallucinations, fluent but factually incorrect statements, remain a persistent risk despite technical progress, particularly when models are queried about recent evidence, rare conditions or fine-grained numerical reasoning[92]. These errors are especially hazardous in clinical contexts where a confident-sounding incorrect assertion may go unchallenged.

A deeper problem concerns the nature of LLM output itself. When a clinician asks an LLM for a diagnosis or risk estimate, the model generates a sequence of tokens, each chosen based on its likelihood given the preceding text, not a calibrated estimate of disease probability. Rephrasing the question, reordering clinical facts or changing terminology can shift the model's apparent 'confidence' even when the underlying evidence is unchanged. Token-level output is therefore not equivalent to a posterior probability over a clinical outcome; treating it as such is a category error rather than a calibration problem. Generative models trained on EHRs face an analogous issue: their reported 'risk scores' reflect frequencies over simulated patient timelines, not validated predictions of clinical events, and they have not been shown to generalize reliably across healthcare systems or patient populations[108].

Beyond these output-level concerns, updating a single LLM's knowledge is resource intensive and technically challenging without guarantee of success[109], so deployed models progressively fall behind evolving treatment guidelines, newly approved therapies and emerging biomarkers. Although modern LLMs support context windows of up to one million tokens, attention is not uniform across that span: retrieval performance degrades for information positioned far from the query, meaning that longitudinal patient records spanning years of clinical visits, neuroimaging and biomarker data cannot be processed in full without some information being effectively lost[110] (Box 2).

https://doi.org/10.1038/s43587-026-01186-z

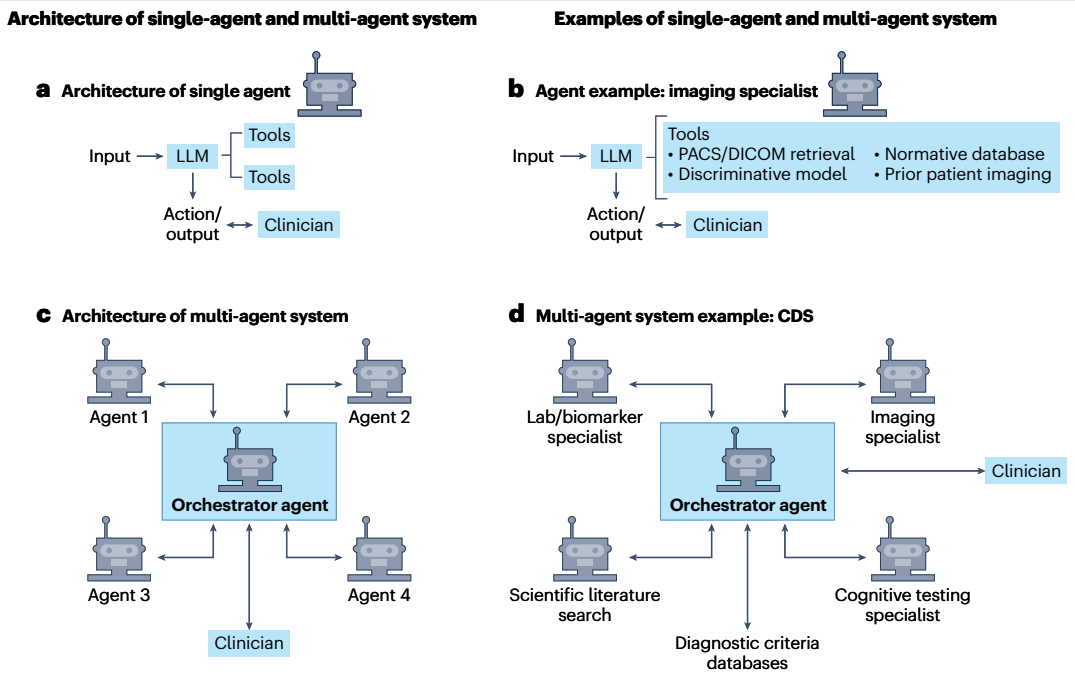
BOX 2

The potential for agentic AI systems to address single LLM limitations

Agentic AI systems are compound architectures in which one or more LLMs orchestrate specialized components including tools such as web search, code execution and data retrieval; curated knowledge bases; and other specialized AI models, to complete complex, multistep tasks[22,23] (Box 2 figure). Unlike a single LLM responding to a prompt in isolation, an agentic system decomposes a clinical problem into steps, selects appropriate tools for each and verifies its own outputs before synthesizing a response grounded in authoritative sources. In a multi-agent configuration, specialist agents (for example, for imaging interpretation, literature retrieval, biomarker contextualization and clinical history synthesis) can run in parallel, with an orchestrating agent integrating their outputs into a unified recommendation. Systems of this kind are demonstrating impressive diagnostic capabilities, with multi-agent diagnostic pipelines and retrieval-augmented CDS now demonstrated in the literature[78,79].

These design choices target the limitations described in Box 1. Jagged intelligence is managed through scaffolded planning, in which the orchestrating LLM breaks tasks into verifiable subtasks and reroutes steps that fail internal validation[22,23,107]. Hallucinations can be partially mitigated by grounding outputs in real-time retrieval, such as from clinical trial registries and AI-powered medical evidence platforms (for example, OpenEvidence, which synthesizes peer-reviewed literature to answer clinical questions), rather than relying on memorized training data[92,111].

Agentic architectures enable richer integration of clinical context, supporting more accurate and nuanced interpretation of patient data and of outputs from machine learning models. A specialized agent can invoke a validated diagnostic or prognostic model, preserve its numerical output exactly (an estimate) and translate it for the clinician by explaining where and when the model applies, how the result should be weighed for this particular patient and how it fits with the rest of the patient's data. The LLM contextualizes and communicates the output; an appropriate, validated machine learning model produces it. Through context engineering, the agent dynamically invokes frameworks, biomarker reference databases, clinical guidelines and the patient's full longitudinal record at the moment of decision, overcoming the knowledge cutoff and context window limitations inherent to static models[109–112].



Box Fig. 2 | Examples of agentic AI architectures. a–d, The panel on the left provides an overview of a single agent architecture (**a**) and a multi-agent architecture (**c**), and the panel on the right provides examples of each (**b** and **d**, respectively). In a single agentic AI system (**a**), an LLM receives structured and unstructured clinical input, invokes specialized tools (for example, diagnostic classifiers, knowledge retrieval, EHR queries) and produces an action or output that is reviewed by the clinician. In this example (**b**), an agent specialized for neuroimaging interpretation retrieves a DICOM image, can use models for atrophy quantification compared to a normative database as well as models for classification (for example, Alzheimer's disease and Lewy body disease) and, if available, can compare this image to that patient's prior images. In a multi-agent system (**c**), multiple LLM-based agents, each with access to their own inputs and tools, work in parallel in a coordinated loop, with one agent orchestrating the others. Running specialist agents simultaneously allows complex clinical tasks to be decomposed and executed concurrently, with results synthesized before the combined output is reviewed by the clinician, preserving human oversight throughout. In this example (**d**), the multi-agent system is for CDS to diagnose a complex clinical case. The clinician interacts iteratively with the orchestrator AI agent.

(continued from previous page)

Effective clinical deployment further requires harness engineering: the deliberate design of the environment, scaffolding and feedback loops within which agents operate, a paradigm already demonstrated at scale by AI coding assistants such as Anthropic's Claude Code[113,114]. Standardized connectors, including FHIR, SMART on FHIR and the Model Context Protocol, extend this reach into clinical infrastructure, allowing agents to securely retrieve EHR data, laboratory results and knowledge bases at the moment of decision[52,53,115].

The potential clinical benefits of these architectures make the difficult work ahead worthwhile. No architecture eliminates the challenges of clinical AI. Realizing these capabilities will require deliberate design, validation, ongoing monitoring and iterative updating of the agentic system as its underlying components evolve. Within this work, constrained autonomy remains crucial: the clinician retains full decision-making authority, reviewing structured, verifiable recommendations before any action is taken[24].

High-quality, standardized data collection

High-quality, standardized data collection is the foundation of our roadmap (Table 1) and should first prioritize what clinicians can interpret and are currently using to diagnose neurodegenerative disease in both the clinic and research settings. In the following sections, we examine the key data types currently integral to ADRD diagnosis and how AI can enhance their collection and interpretation: the patient's history, physical examination, neuropsychological testing, neuroimaging and biomarkers. Agentic AI systems will transform the collection of the patient's history, physical examination and neuropsychological testing, enabling consistency and standardization not currently possible when relying on providers with different training and practice patterns (Box 2 and Fig. 1b). These AI systems can then integrate these data alongside neuroimaging and biomarkers for interpretation.

Gathering the patient's history

Recent advances in LLM capabilities have made comprehensive remote history collection with voice-based conversational agents feasible for the first time. This includes natural conversational interaction that can match or exceed clinician efficiency and expressions of empathy[19], engage with persons living with dementia[25] and adapt to various educational backgrounds and languages[26].

Clinicians, faced with limited time, struggle to gather and document comprehensive patient histories from distressed patients or their caregivers, leading to redundant interviews across clinicians and potentially delaying diagnosis. Moreover, incomplete documentation restricts the valuable clinical data available for AI-driven decision support systems. Voice-based conversational agents powered by LLMs could transform history collection by enabling these interviews to be conducted over the phone outside time-limited clinic visits, increasing healthcare access for patients who are of lower socioeconomic status, have limited English proficiency and/or have low digital literacy[27–30]. This can democratize the highly specialized interviews needed to differentiate between neurodegenerative conditions (for example, syndromes most compatible with Alzheimer's disease, frontotemporal dementia or Lewy body disease) and could be integrated at various points along the patient's care pathway.

There are several considerations for the efficacy and safety of interactions between patients and conversational agents. Interviews will need to be semi-structured as LLMs could engage in irrelevant or dangerous topics. Continuous real-time monitoring systems are needed to analyze the conversation for content relevance and for dangerous content such as expressions of suicidal ideation, automatically triggering human intervention. A hybrid model that combines automated conversational agents with timely clinician oversight and decision-making is strongly advocated. Bias will also need to be identified and mitigated (for example, these tools can have worse performance with non-English-speaking people)[31].

Most clinicians would not realistically adopt a system requiring them to read a lengthy transcript, but LLM summarization capabilities enable workflow integration by converting these interviews into a history of present illness[32]. Reliable summarization is especially critical in cognitive impairment contexts, where accurate documentation of subtle symptom changes over time is essential for early diagnosis and management.

Neuropsychological testing

When concerning symptoms are identified, a cognitive assessment with objective measures of cognitive impairment is required. Cognitive assessments should be scalable, without requiring highly trained staff, and designed to identify relative weaknesses of specific cognitive and behavioral domains to enable identification of patterns helpful with diagnosis, going beyond brief screening assessments. Traditional assessments are time-consuming, require specialized training and often capture only summary scores rather than the detailed, domain-specific performance data valuable for AI interpretation.

Digital cognitive assessments will be necessary to enable a comprehensive AI system for ADRD, and they address the above limitations by capturing nuances in performance (trial-level data, response times and error patterns) unavailable with paper-based tests. Their ability to integrate directly into the clinician's workflow within the EHR has driven adoption, as shown with the TABCat-BHA[33]. The digital cognitive assessment landscape is advancing rapidly[34,35], with clinically available tools continually improving in accuracy, accessibility and ecological validity[34,35]. Virtual reality assessments represent one particularly promising direction[36], with emerging LLM integration enabling virtual psychometrists that could fundamentally transform how cognitive evaluations are conducted.

Physical examination

Physical examination continues to be important in obtaining a diagnosis and the motor findings associated with Lewy body disease, corticobasal syndrome, progressive supranuclear palsy, amyotrophic lateral sclerosis, cerebrovascular disease and many other conditions. In addition, the absence of motor findings in Alzheimer's disease is also important. Although an in-person physical examination is ideal, physical examination findings for these syndromes can often be uncovered through video visits, which can increase accessibility for patients who cannot feasibly travel to see a specialist. Capabilities with promising results in the research setting include video interpretation of speech and speech patterns, motor planning and execution, gait, balance and others[37]. Notable examples include recent studies showing short videos captured on consumer-grade technology can detect parkinsonism[38,39].

Neuroimaging and biofluid biomarker data

Neuroimaging with magnetic resonance imaging (MRI) and positron emission tomography (PET) is a diagnostic cornerstone, aiding in exclusion of treatable causes of dementia, assessing vascular etiologies, and observing atrophy patterns (MRI) or biomarker location (PET). MRI is becoming more accessible around the country with several creative solutions[40].

Biomarkers are now a valuable component of the diagnostic process. Although amyloid PET and cerebrospinal fluid (CSF) biomarkers for Alzheimer's disease have not been easily scalable or accessible, multiple plasma-based biomarkers are becoming available in the clinic setting[41–43]. Assays of CSF and, more recently, skin biopsy samples are

https://doi.org/10.1038/s43587-026-01186-z

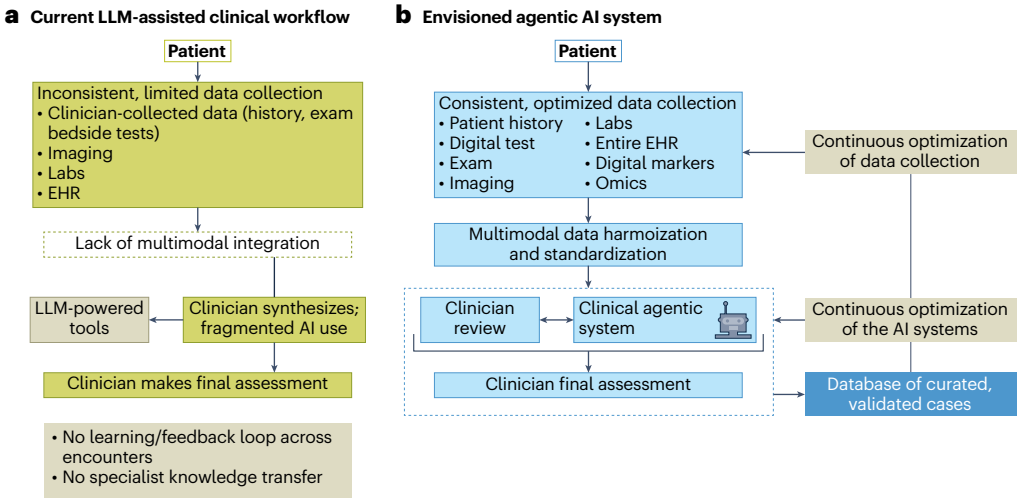


Fig. 1 | Current state-of-the-art LLM-assisted clinical workflow and the envisioned agentic AI system enabling a learning healthcare system. **a**, Clinician-collected data, gathered during time-limited encounters, vary in scope depending on available time, training and practice setting, and are not integrated with other data before reaching any tool. Current LLM-powered tools are unlocking previously unusable unstructured data, but they are limited by what the clinician has time to provide, are unimodal and do not work in coordination. The clinician interacts with each tool separately and synthesizes outputs, dividing visit time between data collection, tool interaction and shared decision-making. No learning loop exists across encounters, and specialist knowledge does not transfer between clinicians. **b**, The envisioned agentic AI system (blue) consistently collects data independent of visit-time constraints, including through conversational agents, specialized testing through digital platforms and physical examination through video-based analysis. These inputs flow through a harmonization pipeline into a standardized multimodal dataset, the foundation of system quality. This integration layer, with agentic synthesis of each data element in full patient context, also enables data otherwise infeasible for clinicians to leverage directly (for example, digital biomarkers, omics and the entire EHR). The dataset is accessible in parallel to the clinician and a clinical agentic system whose specialized agents provide assistance including documentation, in-basket management and clinical decision support; the latter coordinates discriminative models for imaging interpretation, laboratory contextualization and retrieval of updated medical knowledge and guidelines. The clinician retains full decision-making authority, reviewing data and AI-generated recommendations with the patient. The visit, including nuanced symptoms, is captured by ambient scribe, reintegrated with the multimodal dataset, and provided to the clinical agentic system for further recommendations. Feedback loops enable continuous learning across encounters and clinicians: feedback informs data collection, and curated, validated cases inform ongoing improvement of the clinical agentic system.

promising for the detection of α-synuclein, which is a biomarker of Parkinson's disease, dementia with Lewy bodies and multiple-system atrophy[44–46]. The implementation and real-world accuracy of these biomarkers is highly promising[43]. Experts in the field have emphasized that accurate interpretation of these biomarkers still requires clinical context[47,48], which is a concept that can be extended to any data type. Therefore, an AI system must integrate data across modalities with unique patient characteristics to inform diagnosis and management.

Multimodal data integration: creating a unified clinical picture

Standardized data collection is the essential first step towards a timely diagnosis, enabling both personalized management plans and meaningful research contributions. Methods must be easy for patients and their caregivers, as well as clinicians, as this can increase access to care and ensure data are from diverse patient populations and clinic settings. However, achieving global standardization at the point of care is extremely challenging in diverse, resource-constrained settings. Therefore, we advocate for an 'AI-first' data infrastructure approach[49]. Rather than imposing rigid collection protocols on frontline clinicians, this approach prioritizes backend curation and harmonization layers that can transform heterogeneous real-world clinical data into AI-ready formats. Standardized, high-quality, data conversion for each modality not only facilitates efficient storage and sharing but also underpins the continuous development of effective AI systems. The quality of an agentic AI system is dependent on the quality of the data that it learns from, making this phase the foundation of our roadmap (see Fig. 1b, 'Consistent, optimized data collection' and 'Multimodal data harmonization and standardization').

Both existing and future datasets require proper curation according to FAIR principles (findable, accessible, interoperable and reusable), including comprehensive metadata, clear governance frameworks and quality-control procedures[50]. National organizations such as the National Institutes of Health (NIH), CMS, advocacy organizations and other entities could potentially facilitate the development of standards (for example, Digital Imaging and Communications in Medicine (DICOM) for neuroimaging) and monitoring, although there are many other governance models that have been written about elsewhere[51].

The most challenging aspect of a scalable AI system in the US healthcare system will probably be multimodal data integration across diverse and complex EHR systems while maintaining strict compliance and data protection standards according to the Health Insurance Portability and Accountability Act. Practical pathways forward are forming with two decades of progress in interoperability standards, and this momentum has accelerated as AI has dramatically increased the value of data; however, implementation will be difficult across healthcare settings due to varying degrees of resource constraints, technical expertise and competing institutional priorities.

The Substitutable Medical Apps, Reusable Technology (SMART) on Fast Healthcare Interoperability Resources (FHIR) framework enables applications to read and write standardized health data across vendors[52,53], with major EHR systems like Epic and Oracle already supporting robust integration ecosystems[54,55]. Real-world examples demonstrate this capability. UCSF's BRIDGE dashboards integrate multimodal sources across any EHR compatible with SMART on FHIR[56]. The Observational Medical Outcomes Partnership (OMOP) Common Data Model extension for imaging (MI-CDM) was used to convert 1 million DICOM series from an Alzheimer's neuroimaging cohort into OMOP tables alongside clinical EHR data[57]. Similar advances have integrated genomic, wearable and digital cognitive assessment data into the EHR[58,59]. Research initiatives have created environments with standardized, multimodal data from multiple institutions, such as the Alzheimer's Disease Data Initiative (ADDI) and AD Workbench, a

Table 1 | Agentic AI integration roadmap for ADRD care

Phase	Rationale and objectives	Implementation strategies	Potential key stakeholders	Expected impact
(1) High-quality, standardized data collection	**Near term**			
	• Build infrastructure to collect standardized, high-quality, multimodal clinical data • Design tools for data collection to be accessible and easy for diverse patient populations and clinic settings • Establish structured formats for data currently lacking standardization • Leverage LLM ability to capture unstructured, nuanced text	• Encourage and leverage ambient scribes for high-quality documentation • Establish multi-institutional data standards across all modalities	• Professional societies • Research institutions • Clinicians (dementia specialists, primary care providers, neuropsychologists) • Patient advocacy groups • Domain experts for each data type (neuroimaging, biomarkers) • Health informatics specialists	• Comprehensive data collection that reduces patient and clinician burden • Standardized assessment across care settings
	Long term			
	• Proper curation according to FAIR principles	• Create repository of standardized, multimodal ADRD datasets accessible in the clinic setting • Validate implementation and effectiveness of AI-enabled data collection in diverse settings (digital cognitive assessments, voice-based conversational agents for patient history over the phone, video-based analysis of physical examinations) • Integrate passive monitoring from digital biomarkers in clinical care	• National organizations (NIH, CMS)	• Comprehensive multimodal data available at point of care across settings • Data prioritization across institutions evolves with the latest evidence (for example, implementation of more accurate, evidence-based biomarkers)
(2) AI CDS	**Near term**			
	• Enable explainable diagnostic reasoning that clinicians can verify and trust, grounded in the latest evidence base • Support personalized management decisions including complex therapy eligibility criteria • Augment clinician capabilities through clear presentation and interpretation of complex multimodal data using discriminative and generative models	• Train models to interpret ADRD-specific multimodal data • Develop LLM-based systems capable of explaining decision support recommendations mimicking clinician decision-making • Use methods such as retrieval augmented generation (RAG) to ground decision support in latest evidence base • Pioneer/larger health systems piloting CDS tools in specialty clinics using multimodal data integration	• AI researchers • Clinicians (dementia specialists, primary care providers, neuropsychologists) • Health equity researchers • Implementation scientists • Payers/insurers • Healthcare economists • Policymakers	• Interpretable decision support • Personalized care recommendations • Increased identification of clinical trial candidates
	Long term			
	• Enable decision support that approaches specialist-level assessment across care settings.	• Validate implementation and effectiveness of AI decision support across diverse clinical settings and patient populations served, from academic specialty centers to resource-limited primary care practices		• Increase early and accurate diagnosis in primary care and specialty settings, decreasing diagnostic disparities • Expansion of multimodal CDS tools to lower-resourced health systems
(3) Clinical integration and workflow	**Near term**			
	• Address barriers to health technology adoption in clinical settings • Alleviate clinician burnout, designing approaches that enhance rather than disrupt existing workflows • Avoid patient/caregiver burden • Maintain meaningful clinician oversight while maximizing automation benefits	• Study clinician–AI interaction to optimize clinical utility/safety and implementation in both specialty and primary care settings (mitigating underutilization and overreliance) • Build EHR-compatible tools reducing documentation burden and clicks • Create patient–AI interfaces that enhance accessibility and reduce burden • Develop implementation metrics that assess clinical value and workflow efficiency	• Health system leaders • Clinical operations staff • Frontline clinicians • EHR vendors • Implementation scientists • Payers/insurers • Healthcare economists • Policymakers	• Time savings allowing more meaningful patient–provider interactions • Enhanced clinical decision-making with evidence-based care becoming the path of least resistance • Improved clinician satisfaction and retention • AI systems with practical clinician oversight • Implementation evidence from early-adopter sites informing broader rollout
	Long term			
		• Develop AI assistants that can adapt interaction/explanations to clinician level of expertise • Build systems for automated follow-up that maintain human oversight • Implement training programs to optimize clinician–AI collaboration		• Tools easily implemented in any care setting regardless of available resources • Tools implemented in an economically sustainable manner

https://doi.org/10.1038/s43587-026-01186-z

Table 1 (continued) | Agentic AI integration roadmap for ADRD care

Phase	Rationale and objectives	Implementation strategies	Potential key stakeholders	Expected impact
(4) Validation and monitoring	**Near term**			
	• Ensure patient safety • Prevent model drift and degradation • Identify and mitigate biases • Inform regulatory framework	• Establish validation protocols using the FAVES framework • Create monitoring dashboards for model performance using benchmarks and specialist assessments	• National organizations • Professional societies • Dementia specialists and neuropsychologists • Regulatory bodies • Patient advocacy groups • Model governance bodies • Policymakers	• Regulatory approval pathways for AI tools • Early detection of model performance degradation • Technical quality assurance standards for clinical AI
	Long term			
	• Ensure cost-effectiveness and high-value care • Establish sustainable governance structures	• Develop neuropathology and biomarker correlation where feasible • Establish independent oversight committees		• Consistent AI performance across diverse clinical and geographic settings
(5) Continuous learning and improvement	**Near term**			
	• Systematically integrate clinician feedback, patient outcomes and emerging research to refine models while preventing quality degradation • Ensure models continue to address evolving real-world needs • Ensure only high-quality data are used for model updates • Ensure community engagement	• Establish passive and voluntarily active feedback mechanisms for clinicians, and develop protocols for curating feedback • Create patient outcome tracking systems • Build real-time literature monitoring systems to update the RAG database with human oversight • Establish community scientific partnership boards	• Clinical research community • Individuals living with dementia and their families and caregivers • Model governance bodies • Health informatics specialists	• Rapid incorporation of new evidence • Adaptation to evolving clinical practices
	Long term			
		• Implement semi-open models with robust governance, including independent oversight and model update protocols • Create clinician communities for model improvement		• Models that improve with use • Knowledge sharing across institutions
(6) Ethics and risk management	**Near term**			
	• Adopt and advance ethical guidelines to identify and mitigate patient harm as the greatest risk • Ensure fairness, privacy, explainability and clear accountability • Protect healthcare workers' rights as well	• Create efficient adverse event reporting systems • Establish robust privacy safeguards • Ensure transparency in model limitations and potential biases • Design redundant safety systems	• Legal advisors • Professional societies • Patient advocacy groups • Healthcare ethicists • Policymakers	• Accountability mechanisms that assign appropriate responsibility among providers, developers and institutions • Protected patient autonomy • Transparent AI deployment • Established remediation pathways when harm occurs
	Long term			
	• Ensure sustained equity as AI systems scale	• Develop ADRD-specific health AI ethics committee • Create AI ethics certification standards for ADRD clinical applications • Establish liability and indemnification frameworks with legal and regulatory input	• ADRD-specific ethics committee	• Equitable access to AI-enhanced care regardless of geography, resources or demographics

This roadmap conceptualizes a continuous learning healthcare system where insights from monitoring and implementation (phases 3–6) feed back into improving data collection standards and model development (phases 1 and 2), creating an ongoing cycle of improvement.

https://doi.org/10.1038/s43587-026-01186-z

platform with the compute resources necessary for multi-institutional collaboration[60]. The Global Research and Imaging Platform (GRIP), a secure, cloud-based open-access ecosystem, extends this model by providing modular workflows, tools for integrating and analyzing complex multimodal datasets, and secure spaces for collaborative, community-driven development. In parallel, JupyterHealth, an open, modular software platform for health data ingestion and management, spans the full life cycle of digital health workflows from data ingestion, data standardization and transformation, secure storage and authentication to data management and analysis within clinical workflows. Building on HL7 FHIR, Open mHealth and related interoperability standards, these initiatives show how open, community-driven infrastructure can be adopted by diverse health systems, researchers and clinicians.

LLMs are likely to accelerate interoperability further. Recent work demonstrated an LLM translating free-text EHR notes into FHIR with impressive accuracy, suggesting that LLMs may enable interoperability driven less by rigid schemas and more by flexible 'semantic standards', dynamically reshaping data into whatever structure downstream systems require[61].

Collectively, these advances demonstrate that multimodal EHR data integration is possible at institutional scales. However, current implementations remain limited to highly motivated, well-resourced institutions, with widespread adoption across diverse healthcare settings not yet realistic. Nevertheless, rapid policy and technical advances in interoperability create urgency to pilot AI systems integrating multimodal data now, positioning healthcare systems for broader deployment as these capabilities mature. Furthermore, to scale these solutions beyond research cohorts into routine care, we must address barriers that are not solely technical, but largely issues of governance. Divergent policies on data access and privacy across institutions create silos that hinder the training of robust AI models. An AI-ready framework requires not only technical interoperability (for example, handling different data formats) but also semantic alignment and dynamic governance protocols that can manage consent and privacy across borders and institutions.

Enhancing the human element and laying the foundation for AI decision support

We envision that AI-driven data collection in a consistent and standardized manner (that is, history, cognitive assessments and physical examinations) will serve as a foundation for reducing errors and enhancing clinician–patient connection by shifting clinic visit time from data collection/documentation to more meaningful tasks: reviewing findings with patients, answering questions and engaging in shared decision-making about care preferences and values. These data can be presented alongside already standardized neuroimaging and biomarkers in an explainable manner.

This infrastructure enables both immediate diagnostic capabilities and continuous learning, which will allow the AI system to adapt as diagnostic tools evolve (for example, streamlining patient history as biomarkers become more accurate). Practically, implementation will probably begin with larger healthcare systems capable of multimodal data sharing. Smaller systems will require demonstrated value before participating in voluntary data exchange programs such as the Trusted Exchange Framework and Common Agreement (TEFCA), a government initiative designed to facilitate nationwide health information exchange. With sustained effort and evidence of benefit, we can realize the full potential of an AI system serving as a decision support tool in ADRD care.

AI CDS

High-quality data collection and standardization outlined above enables both the clinician and the AI CDS system to efficiently access, integrate and interpret multimodal patient data, avoiding the risk of false positives from relying on one test result (Table 1, phase 2). Our perspective on AI decision support centers on augmenting rather than replacing clinician capabilities by highlighting important diagnostic features from this complex multimodal data, explaining its reasoning transparently[17,20,62]. In this way, AI can be a collaborative partner, rather than a passive tool.

AI research in ADRD decision support

Despite successes with disease classification tasks and their interpretability, traditional machine learning approaches (for example, logistic regression, support vector machines and random forests) face considerable challenges with the multimodal, heterogeneous datasets in ADRD research. These methods require extensive manual feature engineering to extract structured representations from diverse data types, ranging from MRI and PET scans to tabular EHRs and unstructured clinical notes. This introduces potential biases and information loss as well as limits scalability and adaptability. More advanced computational models are needed that can process and integrate relational, unstructured and multimodal data to capture the nuanced, multifactorial nature of cognitive decline.

Deep learning architectures, particularly convolutional neural networks, have enabled remarkable precision in identifying disease-specific features from neuroimaging[63], such as atrophy patterns or cortical thinning[64]. Recent multimodal AI models incorporating clinical, laboratory and imaging data hold the most promise[65–67], with recent successful multimodal models being able to differentiate neurodegenerative subtypes or stages of disease with high accuracy. Multimodal fusion, combining neuroimaging, neuropsychological assessments, functional evaluations, fluid biomarkers and structured data from clinical history, has driven successive advances in ADRD classification: from convolutional neural network-based models matching neurologist-level differential diagnosis accuracy across eight cohorts[68], to transformer architectures distinguishing ten dementia etiologies across over 50,000 participants with AI-augmented assessments outperforming clinician-only evaluations by 26%[65], to frameworks estimating regional amyloid and tau burden from standard clinical data with postmortem neuropathological validation[69]. Although commercial solutions exist for AI-based regional normalized brain volumetry, it remains unclear how much such methods are affected by image acquisition parameters and how such data are best incorporated into a diagnosis for an individual patient[70–72].

Despite advances, deep learning models face substantial implementation challenges. These include substantial data requirements, overfitting risk with small ADRD datasets and sometimes limited interpretability[73]. This lack of transparency complicates clinical adoption by undermining provider trust, highlighting the need for robust validation protocols, diverse training datasets and user interfaces that translate complex model inferences into clinically meaningful explanations[74].

Using agentic AI for ADRD decision support

Recent advancements in LLMs may substantially enhance ADRD care and research, offering new approaches to clinical data interpretation and diagnostic uncertainty reduction. Unlike traditional discriminative models constrained by rigid classification frameworks, LLMs process extensive natural-language inputs, including clinical documentation and patient-generated speech, while generating probabilistic assessments from incomplete or ambiguous clinical data. They have demonstrated impressive medical knowledge that could serve as CDS. For example, GPT-4 exceeded mean human performance on a neurology board-style examination[75] and provided neuropathologic differential diagnoses similar to those of experts[76].

Successful integration of AI in ADRD necessitates a complementary approach using both discriminative models for distinguishing well-defined populations and generative models for managing diagnostic uncertainty and hypothesis generation. This combination can

be accomplished through an agentic AI system, which enhances both the accuracy and generalizability of decision support tools.

Agentic AI systems can address a fundamental healthcare challenge: delivering the right information to the right patient at the right time. In practice, the clinician could review AI-generated insights about the patient before the appointment, including a differential diagnosis with explanations and then provide recommendations for next steps in workup and management based on the latest evidence-based guidelines. This need is growing with the complexity of emerging therapies, particularly amyloid-targeting treatments with numerous inclusion/exclusion criteria and safety requirements making patient selection difficult even for specialists. AI systems can systematically evaluate these complex criteria, reducing errors and helping to identify suitable candidates for specific interventions earlier. Such systems could also suggest personalized interventions ranging from disease-modifying therapies to supportive care like psychotherapy, physical therapy and cognitive maintenance strategies including lifestyle modifications, as well as matching patients with clinical trials. During appointments, clinicians verify findings with patients, guide shared decision-making about next steps and leverage AI to provide explanations optimized for that patient's level of education, age and culture[77].

A practical advantage of agentic AI in ADRD care is its ability to integrate multimodal clinical data into a single, clinician-interpretable narrative. These systems can combine text (clinic notes, radiology reports), structured data (labs, vitals, medications) and higher-dimensional signals (for example, imaging, speech, actigraphy-based wearables) in two main ways. One approach is modality-to-language standardization where imaging features, test results and other non-text input features are converted into textual or tabular summaries that can be consumed by LLM-based agents. Gallingani et al.[78] developed an agentic AI system that simulates a multidisciplinary clinical team. This system achieved high accuracy in classifying pathology-confirmed primary progressive aphasia variants using multimodal inputs, including clinical notes, neuropsychological and language assessments, and MRI-derived atrophy scores. The system preprocesses these data, extracting features that are then converted into textual summaries and then uses specialized agents (for behavioral neurology, neuropsychology and neuroimaging), whose analyses are synthesized and reviewed by a set of prompt-engineered specialized AI agents. Importantly, the output report is designed to be interpretable in familiar clinical terms that resemble how clinicians present cases, allowing clinicians to evaluate the evidence supporting the differential and the predicted underlying pathology. Microsoft's MAI-DxO (MAI Diagnostic Orchestrator) illustrates a related orchestration pattern: sequential diagnostic reasoning that surfaces clinician-readable intermediate steps and culminates in an auditable natural-language synthesis of what was asked, what was observed and why a recommendation follows[79]. A second approach is to build on multimodal foundation models that reason jointly over vision and language, directly incorporating images or documents alongside text during dialog and inference while maintaining a clinician-facing summary for review, as demonstrated by multimodal AMIE (Articulate Medical Intelligence Explorer), an agentic AI system developed by Google for telehealth-style consultations[80]. Together, these examples illustrate emerging agentic AI system architectures that support multimodal data integration and clinician-facing decision support.

Clinical integration and workflow

Clinical integration and workflow principles (Table 1, phase 3) are foundational to every phase of this roadmap. The principles of enhancing rather than disrupting clinician workflow, minimizing patient and caregiver burden and making evidence-based care the path of least resistance shape how data are collected (phase 1), how decision support is designed to augment clinician capability (phase 2), how clinician–AI interaction patterns are monitored (phase 4) and how the system continuously learns from routine care without adding clinician burden (phase 5). Here we focus on the question that determines whether any of these design choices can be realized at scale: economic viability and the near-term pathways that can bring agentic AI into routine clinical practice.

Economic considerations

Economic viability of healthcare AI systems remains uncertain, with substantial unknowns around return on investment, infrastructure costs and reimbursement pathways. A recent systematic review demonstrated AI cost-effectiveness across clinical domains from a healthcare system perspective through quality-adjusted life-years gained, disability-adjusted life-years averted and diagnostic yield improvements, reducing unnecessary studies and treatments[81]. Other studies have found earlier diagnosis to be cost-effective for Alzheimer's disease, particularly when accounting for caregiver economic burden and slowing progression with new disease-modifying therapies[82]. However, there is still some uncertainty as many studies fail to fully account for infrastructure investments, integration expenses and maintenance costs. Further, while the return on investment using AI appears to be favorable for radiology[83], the returns on investment for other applications are mainly estimates based on little data.

There are trends that offer cautious optimism. LLM inference and hardware costs are overall declining, and energy efficiency is improving[84]. Evidence suggests that value-based payment models could incentivize the adoption of AI, although currently reimbursement for AI is through per-use payments[85]. There is national government momentum in enabling new reimbursement pathways as demonstrated by a recent bipartisan bill introduced in April 2025 (The Health Tech Investment Act, or S. 1399), which proposes systematic Medicare coverage for newly Food and Drug Administration (FDA)-authorized AI-enabled medical devices to promote innovation. Healthcare systems have already shown a willingness to invest in AI without clear reimbursement or a concrete financial return on investment as demonstrated by the rapid adoption of ambient scribes, with evidence of decreasing clinician burnout[86]. Although a shared AI system could provide economies of scale among multiple healthcare systems, government funding would probably be needed to incentivize the upfront costs particularly in smaller healthcare systems, and more reliable reimbursement may be needed to cover ongoing maintenance costs.

Electronic consultations (e-consults) offer the most practical near-term implementation pathway, which are asynchronous, provider-to-provider communications that allow clinicians such as primary care providers to seek specialist input through a shared EHR or web-based platform, often resulting in fewer in-person specialist visits[87]. In our experience, e-consults for cognitive neurology typically take less than 10 min to complete. However, due to the limited data available with these consults, many patients still require a long wait for in-person visits exceeding 60 min. The clinical agentic system proposed here could provide the data and decision support necessary for specialists to definitively address more cases in a fraction of the time required for a new patient visit, reducing wait times and patient travel burden especially for rural populations. Furthermore, specialists performing these reviews would simultaneously validate AI system performance, creating a sustainable monitoring mechanism without additional time investment. Demonstrating economic sustainability of these systems will be essential before widespread implementation.

Validation and monitoring

Validation and monitoring (Table 1; phase 4) are essential to ensure that agentic AI systems perform safely and equitably in real-world use. Validation requires comparison with appropriately defined 'gold standards' in a generalizable clinical population, with sufficient statistical power to demonstrate efficacy on a patient-by-patient (not simply group-wise) basis. The vast majority of generative AI diagnostic studies rely on simulated rather than real-world data, highlighting the need for

https://doi.org/10.1038/s43587-026-01186-z

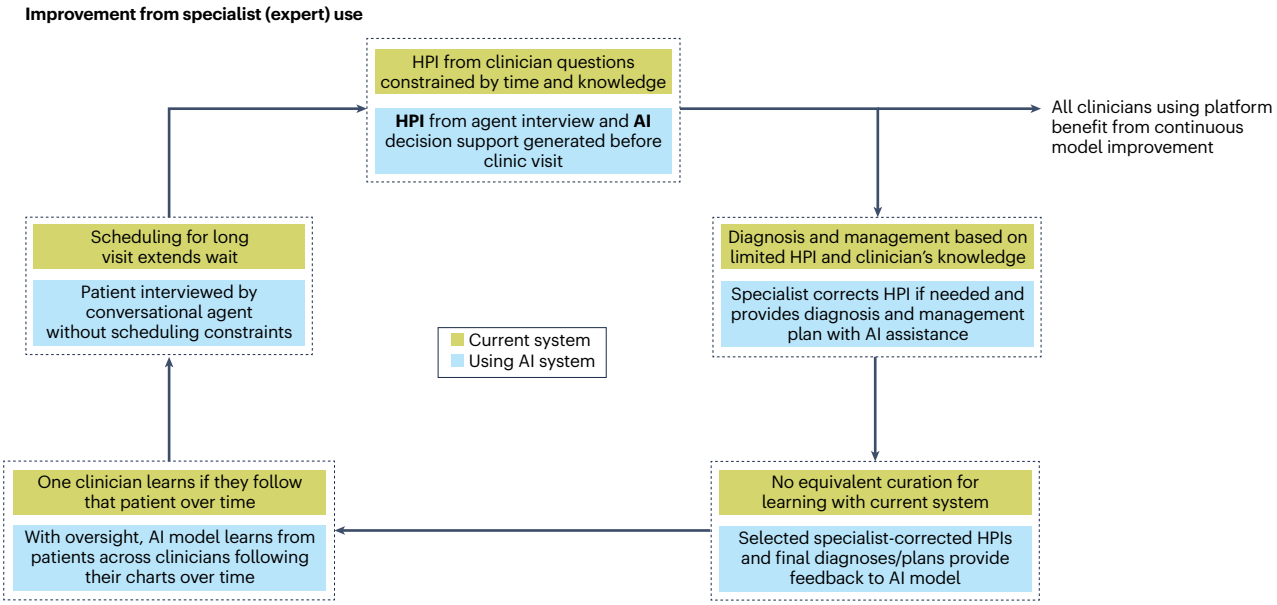


Fig. 2 | Continuous improvement from specialist use. Segments represent steps in the clinical workflow for ADRD. The green color depicts current practice, in which specialist time constraints, long wait times and lack of curation limit learning across clinicians. The blue color depicts the same workflow supported by an agentic AI system: the patient is interviewed by a conversational agent without scheduling constraints, AI-generated decision support is available before the clinic visit based on a comprehensive interview unconstrained by time, the specialist corrects the AI-generated history of present illness (HPI) and provides diagnosis and management with AI assistance, and selected specialist-corrected cases feed back into the model. Model improvements benefit all clinicians using the platform.

more systematic clinical validation[88]. Before real-world implementation, the possibility of misdiagnosis or adverse events must be virtually zero. Most clinical research with the gold-standard neuropathologic verification is performed on well-educated relatively affluent patients with health insurance, often in academic settings. However, in the real world, patients have less education and insurance, greater comorbidities and may be less fluent in English. Therefore, we propose models should also be developed using diverse specialist-confirmed cases in addition to neuropathology. This approach acknowledges both the gold-standard status of neuropathology and the practical need for representative validation.

However, specialist diagnoses contain uncertainty and there is variability even within a single practitioner, with greater variability across different clinicians and institutions. Careful case selection is required for reliable validation datasets in the absence of neuropathological confirmation, for example, prioritizing cases confirmed through highly accurate biomarkers with matching syndromes (for example, Alzheimer's disease biomarkers), clearly meeting criteria known to be highly specific for a certain pathology (for example, progressive supranuclear palsy Richardson syndrome), multiple specialist agreement or longitudinal confirmation of diagnosis through disease progression patterns. This step is crucial to reduce the risk of contaminated model updates.

Ongoing monitoring is essential to prevent model degradation—the unintended decline in model performance, accuracy or safety from improper updates, contaminated training data or inconsistent validation practices. Agreed upon benchmarks must ensure models are fair, appropriate, valid, effective, and safe (FAVES)[89]. Recent initiatives like OpenAI's HealthBench and Stanford's MedArena demonstrate promising approaches for systematic medical AI evaluation across diverse clinical scenarios. These benchmarking frameworks could be adapted for neurodegenerative disease-specific applications, continuously tested on the unique diagnostic challenges these conditions present.

Specialists are best positioned to identify model errors. Much of this monitoring can occur naturally through clinical agentic system use. Specialists could use this system to improve their own efficiency and quality in patient visits and e-consults, positioning them to detect errors in routine practice. Although specialist time is limited, their investment in developing clinical agentic systems scales their expertise enabling primary care providers and general neurologists to definitively manage more cases, reducing unnecessary referrals and increasing their capacity. This represents an evolution in specialist roles, where expertise is leveraged by optimizing agentic systems rather than being confined to individual patient encounters. If cost-effectiveness evidence builds, reimbursement mechanisms could justify dedicated monitoring activities such as batch evaluation of flagged cases. Beyond technical performance, monitoring must evaluate physician–AI interaction patterns, as clinicians often over-rely on or underuse AI capabilities[90,91], informing both system design and clinician training. In real-world settings, it is equally important to monitor how AI use affects clinician efficiency, productivity and job satisfaction.

To ensure coherence across this learning healthcare system, the oversight body governing standards would ideally also oversee validation, monitoring and continuous learning processes discussed in the next section (see 'clinical agent system outputs' and 'feedback loops' in Fig. 1b). While validation and monitoring focuses on quality control and preventing harm, and continuous learning aims to identify opportunities for improvement, these functions are inseparable: continuous learning without rigorous validation risks model degradation from contaminated or unverified updates, and validation without continuous improvement fails to leverage clinical experience. The careful case selection criteria described above apply equally to learning datasets. A unified governance approach ensures consistent standards, maintains institutional knowledge about agentic system behavior and aligns system evolution with healthcare priorities rather than commercial interests.

Continuous learning and improvement

While validation and monitoring prevent harm, continuous learning (Table 1, phase 5) ensures this clinical agentic system advances with emerging knowledge and clinical experience. This requires staying up to date with the rapidly evolving medical literature as well as learning

from carefully validated patient cases in a supervised manner, guided collaboratively by the healthcare community.

Clinician learning through patient cases is currently limited as they do not often see what happens to a patient if that patient does not follow up with them over time. Furthermore, what one clinician learns is difficult to transmit to others, typically requiring case reports or registry submissions that add workflow burden and capture limited information at one point in time. An agentic AI system can address both limitations without adding clinician workload, transforming the clinical workflow cycle as specialist-corrected cases continuously improve model performance for all users (Fig. 2). This is newly feasible because LLMs enable both more complete data capture during routine care (phase 1) and the ability to derive meaning and therefore learn from the resulting unstructured, nuanced clinical data. Feedback loops can rely on passive signal capture by observing whether and how the clinician's decisions diverge from AI recommendations, with optional active feedback available for more detailed input. Built-in processes enable the agentic system to revisit each patient's chart over time for updates, such as an updated diagnosis or treatment response. Clinicians could receive these updates as well, informing their own practice and taking action when needed. Specialist use of the system, discussed above as a monitoring mechanism, simultaneously generates high-value learning on cases where their expertise is most informative. The careful case selection principles from the 'Validation and monitoring' section apply to continuous learning: only high-quality, validated cases from diverse clinical settings should influence model evolution, with ongoing oversight preventing contaminated updates.

Transparency in AI development can be accomplished through using open-source/open-weights models with robust governance. When proprietary models are updated to optimize general use cases such as writing or coding, performance on specific use cases such as healthcare can degrade unpredictably, requiring costly and frequent revalidation cycles, with many models sunsetting every 1–2 years. Open-source models offer full control over updates, enable complete fine-tuning and allow mechanistic interpretability analysis (that is, understanding how an AI model is making decisions). LLMs of smaller size (8 billion–70 billion parameters) can be trained to match or surpass the performance of large proprietary models for more specific, narrower tasks including the task of answering medical questions[92–94], making subspecialty-focused models feasible. Although the computational overhead of these models once made routine clinical deployment impractical, recent parameter-efficient LLM compression methods (for example, 4-bit QLoRA[95] or aggressive edge-device quantization) now enable low-latency, on-device inference suitable for real-world clinical workflows.

Phases 1–5 will probably evolve with near-term deployments that lay a foundation to a long-term learning healthcare system (Table 1), guided by the ethics and risk management discussed in the next section.

Ethics and risk management

Ethics and risk management (Table 1, phase 6) must be addressed and updated as this AI system evolves. Ethical challenges around the use of AI in ADRD have been discussed elsewhere[96]; therefore, a brief overview is provided here with updates that apply across healthcare. Populations with ADRD face unique ethical considerations beyond general AI concerns. As a vulnerable population with declining capacity, autonomy changes over time, requiring informed consent processes that evolve dynamically to involve both patients and caregivers appropriately. This shifting autonomy creates inherent tension between safety monitoring and preserving patient independence and dignity. Design considerations risk amplifying disparities if cognitive decline disproportionately impairs patients' ability to use AI tools, while privacy concerns intensify given the sensitive cognitive and behavioral data collected from individuals who may lack the capacity to understand data usage.

Many ethical considerations have intensified around the use of AI in healthcare, particularly since the widespread enthusiasm with generative AI since November 2022. In response, numerous initiatives have developed frameworks and guidelines to ensure AI applications use the FAVES framework, which will at least in part require transparency with technical performance as well as organizational capacity to manage risks associated with deploying this technology[89]. Fairness has been of particular concern as AI has shown numerous biases[97,98], emphasizing the need for training data from diverse groups as well as ongoing monitoring to identify and quickly mitigate biases based upon the analysis of only socioeconomically privileged populations. The advancement in AI capabilities will bring new challenges with patient privacy and informed consent[99,100]. When using clinical notes for example, removing protected health information may be insufficient to protect privacy[101], potentially requiring more stringent data protection and sharing guidelines. Regardless of the quality of the training data, AI tools are not static, in contrast to some pharmaceuticals whose mechanisms of action may be unclear but are safely used with patients. Therefore, explainability and interpretability[102–104] are important components for a human in the loop to potentially identify signs of model deterioration or drift at the point of care. Ethical considerations traditionally designed to protect patients will probably need to be extended to healthcare workers as well with the rapid evolution of generative AI models[105]. For example, generative AI's ability to mimic providers' appearance, voice and empathetic conversation will probably be of interest to healthcare employers with an overwhelming patient demand. This dynamic landscape has prompted calls for a network of Health AI Ethics Centers to adapt and respond to these ongoing changes to keep pace in a manner that is often difficult for a central governmental agency[105].

We believe patient harm represents the greatest risk of AI implementation in healthcare. This may occur through privacy breaches, adverse psychological effects from AI interactions (including anxiety, depression or suicidal ideation), misdiagnosis, inappropriate treatment recommendations, medication errors and more. Given the enthusiasm for medical AI and the economic incentives driving rapid adoption, we are concerned these risks are underappreciated. The possibility of widespread harm is not insignificant, particularly with an agentic AI system that is given too much autonomy. Legal liability remains unclear—who bears responsibility if patients suffer or die due to AI errors? Therefore, we emphasize that implementing redundant monitoring systems and establishing clear reporting mechanisms are critical to ensure prompt remediation when adverse events inevitably occur.

The current fragmented governance landscape in healthcare AI is still evolving[106]. Although there has been a marked increase in FDA-approved AI-enabled medical devices, the FDA has limited regulatory authority over many healthcare AI applications (for example, they do not regulate applications that only provide recommendations or information that clinicians do not primarily rely on). Other federal agencies such as CMS have regulatory authority over AI implementation limited to entities participating in their programs. In the absence of comprehensive federal regulation, governance is emerging through state legislation, nonprofit guidance (for example, the Joint Commission and Coalition for Health AI providing national guidance for responsible AI use) and private contracts between developers and healthcare organizations[106]. AI integration in diagnostic workflows will require greater involvement from policymakers and government agencies alongside non-governmental entities. Given this fragmented landscape, the unified oversight approach we propose becomes even more essential: integrating data collection standards, validation protocols, monitoring systems and continuous learning under collaborative governance. This model demonstrates how healthcare specialty communities can establish rigorous governance frameworks that complement and collaborate with evolving governmental regulation,

addressing current gaps while maintaining alignment with evolving regulatory requirements.

Conclusion

The integration of agentic AI systems into clinical care offers transformative potential to address many of the most pressing challenges in this unprecedented time for the field of ADRD. In an era with increasing time constraints and an overwhelming volume of medical knowledge, agentic AI systems can support clinicians by enhancing data collection, improving diagnostic accuracy and personalizing management, ultimately enabling more timely and high-quality patient care. Successful implementation of these technologies will require a rigorous framework that prioritizes inclusivity, transparency and ongoing monitoring to ensure ethical and effective use. Collaboration among clinicians, researchers, policymakers, governmental agencies, industry and patients will be critical to establish standards for data collection, model development and real-world application, ultimately fostering a continuously learning healthcare system.

While this roadmap focuses on ADRD care, its principles apply broadly across neurology and other specialties facing similar challenges. With thoughtful development, agentic AI systems can not only mitigate provider shortages and diagnostic delays but also redefine care delivery, creating a future where technology and human expertise work in harmony to improve patient outcomes and advance medical knowledge.

References

1. Hebert, L. E., Weuve, J., Scherr, P. A. & Evans, D. A. Alzheimer disease in the United States (2010–2050) estimated using the 2010 census. *Neurology* **80**, 1778–1783 (2013).
2. GBD 2016 Neurology Collaborators. Global, regional, and national burden of neurological disorders, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol*. 18, 459–480 (2019).
3. Fang, M. et al. Lifetime risk and projected burden of dementia. *Nat. Med*. https://doi.org/10.1038/s41591-024-03340-9 (2025).
4. Bradford, A., Kunik, M. E., Schulz, P., Williams, S. P. & Singh, H. Missed and delayed diagnosis of dementia in primary care: prevalence and contributing factors. *Alzheimer Dis. Assoc. Disord.* **23**, 306–314 (2009).
5. Kotagal, V. et al. Factors associated with cognitive evaluations in the United States. *Neurology* **84**, 64–71 (2015).
6. Babulal, G. M. et al. Perspectives on ethnic and racial disparities in Alzheimer's disease and related dementias: update and areas of immediate need. *Alzheimers Dement.* **15**, 292–312 (2019).
7. Boustani, M. et al. Implementing a screening and diagnosis program for dementia in primary care. *J. Gen. Intern. Med.* **20**, 572–577 (2005).
8. Bernstein, A. et al. Primary care provider attitudes and practices evaluating and managing patients with neurocognitive disorders. *J. Gen. Intern. Med.* **34**, 1691–1692 (2019).
9. Majersik, J. J. et al. A shortage of neurologists – we must act now: a report from the AAN 2019 Transforming Leaders Program. *Neurology* **96**, 1122–1134 (2021).
10. Satiani, A., Niedermier, J., Satiani, B. & Svendsen, D. P. Projected workforce of psychiatrists in the United States: a population analysis. *Psychiatr. Serv.* **69**, 710–713 (2018).
11. Flaherty, E. & Bartels, S. J. Addressing the community-based geriatric healthcare workforce shortage by leveraging the potential of interprofessional teams. *J. Am. Geriatr. Soc.* **67**, S400–S408 (2019).
12. Liu, J. L. et al. Modeling early detection and geographic variation in health system capacity for Alzheimer's disease-modifying therapies. *RAND Corp.* https://doi.org/10.7249/RRA2643-1 (2024).
13. van Dyck, C. H. et al. Lecanemab in early Alzheimer's disease. *N. Engl. J. Med.* **388**, 9–21 (2023).
14. Sims, J. R. et al. Donanemab in early symptomatic Alzheimer disease: the TRAILBLAZER-ALZ 2 randomized clinical trial. *JAMA* **330**, 512–527 (2023).
15. Guiding an Improved Dementia Experience (GUIDE) Model. *CMS* https://www.cms.gov/priorities/innovation/innovation-models/guide/ (accessed 18 April 2024).
16. Singhal, K. et al. Large language models encode clinical knowledge. *Nature* **620**, 172–180 (2023).
17. Singhal, K. et al. Toward expert-level medical question answering with large language models. *Nat. Med.* https://doi.org/10.1038/s41591-024-03423-7 (2025).
18. Tu, T. et al. Towards generalist biomedical AI. *NEJM AI* **1**, AIoa2300138 (2024).
19. Tu, T. et al. Towards conversational diagnostic artificial intelligence. *Nature* https://doi.org/10.1038/s41586-025-08866-7 (2025).
20. Liu X., et al. A generalist medical language model for disease diagnosis assistance. *Nat. Med.* https://doi.org/10.1038/s41591-024-03416-6 (2025).
21. American Medical Association Center for Digital Health and AI. 2026 Physician Survey on Augmented Intelligence 2026. *ama-assn.org* https://www.ama-assn.org/system/files/physician-ai-sentiment-report.pdf (2026).
22. Gupta, R. et al. The shift from models to compound AI systems. *The Berkeley Artificial Intelligence Research Blog* https://bair.berkeley.edu/blog/2024/02/18/compound-ai-systems/ (2024).
23. Acharya, D. B., Kuppan, K. & Divya, B. Agentic AI: autonomous intelligence for complex goals—a comprehensive survey. *IEEE Access* **13**, 18912–18936 (2025).
24. For trustworthy AI, keep the human in the loop. *Nat Med.* **31**, 3207 (2025).
25. Sheehy, L. et al. Development and initial testing of an artificial intelligence-based virtual reality companion for people living with dementia in long-term care. *J. Clin. Med.* **13**, 5574 (2024).
26. Ray, M. et al. Evaluating a large language model in translating patient instructions to spanish using a standardized framework. *JAMA Pediatr.* **179**, 1026–1033 (2025).
27. Tong, L. et al. Telemedicine and health disparities: Association between patient characteristics and telemedicine, in-person, telephone and message-based care during the COVID-19 pandemic. *IPEM Transl.* **3**, 100010 (2022).
28. Nota, S. P. F. T., Strooker, J. A. & Ring, D. Differences in response rates between mail, e-mail, and telephone follow-up in hand surgery research. *Hand* **9**, 504–510 (2014).
29. Pagán, V. M., McClung, K. S. & Peden, C. J. An observational study of disparities in telemedicine utilization in primary care patients before and during the COVID-19 pandemic. *Telemed. J. E Health* **28**, 1117–1125 (2022).
30. Hinton, L. et al. Practice constraints, behavioral problems, and dementia care: primary care physicians' perspectives. *J. Gen. Intern. Med.* **22**, 1487–1492 (2007).
31. Qu, Y. & Wang, J. Performance and biases of large language models in public opinion simulation. *Humanit. Soc. Sci. Commun.* **11**, 1–13 (2024).
32. Van Veen, D. et al. Adapted large language models can outperform medical experts in clinical text summarization. *Nat. Med.* **30**, 1134–1142 (2024).
33. Breithaupt, A. G. et al. Enhancing early detection of cognitive impairment in primary care with the TabCAT-BHA. *Alzheimers Dement.* **21**, e70437 (2025).
34. Chen, L., Zhen, W. & Peng, D. Research on digital tool in cognitive assessment: a bibliometric analysis. *Front. Psychiatry* https://doi.org/10.3389/fpsyt.2023.1227261 (2023).

35. Belleville, S., LaPlume, A. A. & Purkart, R. Web-based cognitive assessment in older adults: where do we stand?. *Curr. Opin. Neurol.* **36**, 491–497 (2023).
36. Mancuso, V. et al. Systematic review of memory assessment in virtual reality: evaluating convergent and divergent validity with traditional neuropsychological measures. *Front. Hum. Neurosci.* https://doi.org/10.3389/fnhum.2024.1380575 (2024).
37. Breithaupt, A. G. et al. Review of artificial intelligence for clinical use in Alzheimer's disease and related dementias. *Semin. Neurol.* https://doi.org/10.1055/a-2744-9871 (2025).
38. Adnan, T. et al. AI-enabled Parkinson's disease screening using smile videos. *NEJM AI* **2**, AIoa2400950 (2025).
39. Deng, D. et al. Interpretable video-based tracking and quantification of parkinsonism clinical motor states. *NPJ Parkinsons Dis.* **10**, 122 (2024).
40. Zhang, K. et al. Recent advances in compact portable platforms and gradient hardware for brain MRI. *Radiology* **315**, e241904 (2025).
41. Ashton, N. J. et al. Diagnostic accuracy of a plasma phosphorylated tau 217 immunoassay for Alzheimer disease pathology. *JAMA Neurol.* **81**, 255–263 (2024).
42. Janelidze, S. et al. Plasma phosphorylated Tau 217 and Aβ42/40 to predict early brain Aβ accumulation in people without cognitive impairment. *JAMA Neurol.* **81**, 947–957 (2024).
43. Palmqvist, S. et al. Blood biomarkers to detect Alzheimer disease in primary care and secondary care. *JAMA* https://doi.org/10.1001/jama.2024.13855 (2024).
44. Arnold, M. R. et al. α-synuclein seed amplification in CSF and brain from patients with different brain distributions of pathological α-synuclein in the context of co-pathology and non-LBD diagnoses. *Ann. Neurol.* **92**, 650–662 (2022).
45. Gibbons, C. H. et al. Skin biopsy detection of phosphorylated α-synuclein in patients with synucleinopathies. *JAMA* **331**, 1298–1306 (2024).
46. Samudra, N. et al. Clinicopathological correlation of cerebrospinal fluid alpha-synuclein seed amplification assay in a behavioral neurology autopsy cohort. *Alzheimers Dement.* **20**, 3334–3341 (2024).
47. Bouwman, F. H. et al. Clinical application of CSF biomarkers for Alzheimer's disease: from rationale to ratios. *Alzheimers Dement.* **14**, e12314 (2022).
48. Dubois, B. et al. Alzheimer disease as a clinical-biological construct—an International Working Group recommendation. *JAMA Neurol.* https://doi.org/10.1001/jamaneurol.2024.3770 (2024).
49. Jasodanand, V. H., Bellitti, M. & Kolachalama, V. B. An AI-first framework for multimodal data in Alzheimer's disease and related dementias. *Alzheimers Dement.* **21**, e70719 (2025).
50. Wilkinson, M. D. et al. The FAIR Guiding principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 (2016).
51. Shah, N. H. et al. A nationwide network of health ai assurance laboratories. *JAMA* **331**, 245–249 (2024).
52. HL7 International. FHIR R5: fast healthcare interoperability resources. *hl7.org* https://www.hl7.org/fhir/index.html (2023).
53. SMART Health IT. *smarthealthit.org* https://smarthealthit.org/ (accessed 25 October 2025).
54. Oracle Health. Oracle Health Millennium Platform APIs. *oracle.com* https://docs.oracle.com/en/industries/health/millennium-platform-apis/index.html (2025).
55. Epic Systems. Epic's home for data sharing and developer tools. *epic.com* https://open.epic.com/ (2025).
56. Brown, E. G. et al. Enhancing clinical information display to improve patient encounters: human-centered design and evaluation of the Parkinson Disease-BRIDGE platform. *JMIR Hum. Factors* **9**, e33967 (2022).
57. Park, W. Y. et al. Breaking data silos: incorporating the DICOM imaging standard into the OMOP CDM to enable multimodal research. *J. Am. Med. Inform. Assoc.* **32**, 1533–1541 (2025).
58. Lau-Min, K. S. et al. Impact of integrating genomic data into the electronic health record on genetics care delivery. *Genet. Med.* **24**, 2338–2350 (2022).
59. Hoffman, S. L. et al. Comprehensive real time remote monitoring for Parkinson's disease using quantitative DigitoGraphy. *NPJ Parkinsons Dis.* **10**, 137 (2024).
60. McHugh, C. P., Clement, M. H. S. & Phatak, M. AD Workbench: transforming Alzheimer's research with secure, global, and collaborative data sharing and analysis. *Alzheimers Dement.* **21**, e70278 (2025).
61. Brat, G. A., Mandel, J. C. & McDermott, M. B. A. Do we need data standards in the era of large language models?. *NEJM AI* **1**, AIe2400548 (2024).
62. Savage, T., Nayak, A., Gallo, R., Rangan, E. & Chen, J. H. Diagnostic reasoning prompts reveal the potential for large language model interpretability in medicine. *NPJ Digit. Med.* **7**, 20 (2024).
63. Esteva, A. et al. A guide to deep learning in healthcare. *Nat. Med.* **25**, 24–29 (2019).
64. Bron, E. E. et al. Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural MRI: the CADDementia challenge. *Neuroimage* **111**, 562–579 (2015).
65. Xue, C. et al. AI-based differential diagnosis of dementia etiologies on multimodal data. *Nat. Med.* **30**, 2977–2989 (2024).
66. Rauschecker, A. M. et al. Artificial intelligence system approaching neuroradiologist-level differential diagnosis accuracy at brain MRI. *Radiology* **295**, 626–637 (2020).
67. Leming, M. & Im, H. Differential dementia detection from multimodal brain images in a real-world dataset. *Alzheimers Dement.* **21**, e70362 (2025).
68. Qiu, S. et al. Multimodal deep learning for Alzheimer's disease dementia assessment. *Nat. Commun.* **13**, 3404 (2022).
69. Jasodanand, V. H. et al. AI-driven fusion of multimodal data for Alzheimer's disease biomarker assessment. *Nat. Commun.* **16**, 7407 (2025).
70. Tanabe, J. et al. Automated volumetric software in dementia: help or hindrance to the neuroradiologist?. *AJNR Am. J. Neuroradiol.* **45**, 1737–1744 (2024).
71. Haller, S. et al. Basic MR sequence parameters systematically bias automated brain volume estimation. *Neuroradiology* **58**, 1153–1160 (2016).
72. Wittens, M. M. J. et al. Diagnostic performance of automated MRI volumetry by icobrain dm for Alzheimer's disease in a clinical setting: a REMEMBER study. *J. Alzheimers Dis.* **83**, 623–639 (2021).
73. Yu, K. H., Beam, A. L. & Kohane, I. S. Artificial intelligence in healthcare. *Nat. Biomed. Eng.* **2**, 719–731 (2018).
74. Aristidou, A., Jena, R. & Topol, E. J. Bridging the chasm between AI and clinical implementation. *Lancet* **399**, 620 (2022).
75. Schubert, M. C., Wick, W. & Venkataramani, V. Performance of large language models on a neurology board–style examination. *JAMA Netw. Open* **6**, e2346721 (2023).
76. Koga, S., Martin, N. B. & Dickson, D. W. Evaluating the performance of large language models: ChatGPT and Google Bard in generating differential diagnoses in clinicopathological conferences of neurodegenerative disorders. *Brain Pathol.* **34**, e13207 (2023).
77. Ayers, J. W. et al. Comparing Physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA Intern. Med.* **183**, 589–596 (2023).
78. Gallingani, C. et al. Agentic generative artificial intelligence system for classification of pathology-confirmed primary progressive aphasia variants. Preprint at *medRxiv* https://doi.org/10.1101/2025.10.28.25338977 (2025).

79. Saab, K., Park, C. & Strother, T. et al. Advancing conversational diagnostic AI with multimodal reasoning. *Nat. Med.* **32**, 1726–1736 (2026).
80. Saab, K. et al. Advancing conversational diagnostic AI with multimodal reasoning. Preprint at https://arxiv.org/abs/2505.04653 (2025).
81. El Arab, R. A. & Al Moosa, O. A. Systematic review of cost effectiveness and budget impact of artificial intelligence in healthcare. *NPJ Digit. Med.* **8**, 548 (2025).
82. Aye, S., Handels, R., Winblad, B. & Jönsson, L. Optimising Alzheimer's disease diagnosis and treatment: assessing cost-utility of integrating blood biomarkers in clinical practice for disease-modifying treatment. *J. Prev. Alzheimers Dis.* **11**, 928–942 (2024).
83. Bharadwaj, P. et al. Unlocking the value: quantifying the return on investment of hospital artificial intelligence. *J. Am. Coll. Radiol.* **21**, 1677–1685 (2024).
84. Stanford HAI. The 2025 AI Index Report. *hai.stanford.edu* https://hai.stanford.edu/ai-index/2025-ai-index-report/ (accessed 11 October 2025).
85. Parikh, R. B. & Helmchen, L. A. Paying for artificial intelligence in medicine. *NPJ Digit. Med.* **5**, 63 (2022).
86. Peterson Health Technology Institute. Adoption of artificial intelligence in healthcare delivery systems: early applications and impacts. *phti.org* https://phti.org/ai-adoption-early-applications-impacts/ (2025).
87. Tardo, L. et al. Evaluation of the impact of neurology electronic consults (e-consults): experiences of a neurology resident clinic in a safety-net hospital. *J. Telemed. Telecare* **31**, 1270–1277 (2025).
88. Bedi, S. et al. Testing and evaluation of health care applications of large language models: a systematic review. *JAMA* https://doi.org/10.1001/jama.2024.21700 (2024).
89. The Federal Register/FIND. Health data, technology, and interoperability: certification program updates, algorithm transparency, and information sharing. Vol 89, 1192. *proquest.com* https://www.proquest.com/docview/2912085944/citation/14F0AB28E9DD4318PQ/1 (2024).
90. Goh, E. et al. Large language model influence on diagnostic reasoning: a randomized clinical trial. *JAMA Netw. Open* **7**, e2440969 (2024).
91. Jabbour, S. et al. Measuring the impact of AI in the diagnosis of hospitalized patients: a randomized clinical vignette survey study. *JAMA* **330**, 2275–2284 (2023).
92. Tonmoy, S. M. et al. A comprehensive survey of hallucination mitigation techniques in large language models. Preprint at http://arxiv.org/abs/2401.01313 (2024).
93. Bucher, M. J. J. & Martini, M. Fine-tuned 'small' LLMs (still) significantly outperform zero-shot generative ai models in text classification. Preprint at https://arxiv.org/html/2406.08660v1 (2024).
94. Hugging Face. Open Medical-LLM Leaderboard - a Hugging Face Space by openlifescienceai. *huggingface.co* https://huggingface.co/spaces/openlifescienceai/open_medical_llm_leaderboard/ (accessed 13 November 2024).
95. Dettmers, T., Pagnoni, A., Holtzman, A. & Zettlemoyer, L. QLORA: efficient finetuning of quantized LLMs. In *Proc. 37th Int. Conf. on Neural Information Processing Systems. NIPS '23* (eds. Oh, A. et al.) 10088–10115 (Curran Associates Inc., 2023).
96. Abadir, P. M., Battle, A., Walston, J. D. & Chellappa, R. Enhancing care for older adults and dementia patients with large language models: proceedings of the National Institute on Aging—Artificial Intelligence & Technology Collaboratory for Aging Research Symposium. *J. Gerontol. A Biol. Sci. Med. Sci.* **79**, glae176 (2024).
97. Obermeyer, Z., Powers, B., Vogeli, C. & Mullainathan, S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* **366**, 447–453 (2019).
98. Leslie, D., Mazumder, A., Peppin, A., Wolters, M. K. & Hagerty, A. Does 'AI' stand for augmenting inequality in the era of COVID-19 healthcare?. *BMJ* **372**, n304 (2021).
99. Savulescu, J., Giubilini, A., Vandersluis, R. & Mishra, A. Ethics of artificial intelligence in medicine. *Singapore Med. J.* **65**, 150–158 (2024).
100. US Department of Health and Human Services. IRB considerations on the use of artificial intelligence in human subjects research. *hhs.gov* https://www.hhs.gov/ohrp/sachrp-committee/recommendations/irb-considerations-use-artificial-intelligence-human-subjects-research/index.html (2022).
101. Sarkar AR, Chuang YS, Mohammed N, Jiang X. De-identification is not enough: a comparison between de-identified and synthetic clinical notes. Sci Rep. 2024;14(1):29669. https://doi.org/10.1038/s41598-024-81170-y
102. Broniatowski, D. A. Psychological foundations of explainability and interpretability in artificial intelligence. National Institute of Standards and Technology (US); NIST IR 8367. *NIST* https://doi.org/10.6028/NIST.IR.8367 (2021).
103. Tabassi, E. Artificial Intelligence Risk Management Framework (AI RMF 1.0). National Institute of Standards and Technology (US); NIST AI 100-1. *NIST* https://doi.org/10.6028/NIST.AI.100-1 (2023).
104. Phillips, P. J. et al. Four principles of explainable artificial intelligence. National Institute of Standards and Technology (US); NIST IR 8312. *NIST* https://doi.org/10.6028/NIST.IR.8312 (2021).
105. Sim, I. & Cassel, C. The ethics of relational AI — expanding and implementing the Belmont principles. *N. Engl. J. Med.* **391**, 193–196 (2024).
106. Mello, M. M. & Cohen, I. G. Regulation of health and health care artificial intelligence. *JAMA* **333**, 1769–1770 (2025).
107. Dell'Acqua, F. et al. Navigating the jagged technological frontier: field experimental evidence of the effects of AI on knowledge worker productivity and quality. *SSRN Electron J.* https://doi.org/10.2139/ssrn.4573321 (2023).
108. Bedi, S., Fries, J. A. & Shah, N. H. How to interpret 'zero-shot' results from generative EHR models. *Nat. Med.* **32**, 404–406 (2026).
109. Shi, H. et al. Continual learning of large language models: a comprehensive survey. *ACM Comput. Surv.* **58**, 120:1–120:42 (2025).
110. Liu, N. F. et al. Lost in the middle: how language models use long contexts. *Trans. Assoc. Comput. Linguist.* **12**, 157–173 (2024).
111. Gargari, O. K. & Habibi, G. Enhancing medical AI with retrieval-augmented generation: a mini narrative review. *Digit. Health* https://doi.org/10.1177/20552076251337177 (2025).
112. He, J. et al. Retrieval-augmented generation in biomedicine: a survey of technologies, datasets, and clinical applications. Preprint at https://arxiv.org/abs/2505.01146 (2025).
113. Lopopolo, R. Harness engineering: leveraging Codex in an agent-first world. *OpenAI* https://openai.com/index/harness-engineering/ (accessed 27 April 2026).
114. Lu, C. et al. Towards end-to-end automation of AI research. *Nature* **651**, 914–919 (2026).
115. Model Context Protocol. What is the Model Context Protocol (MCP)? *modelcontextprotocol.io* https://modelcontextprotocol.io/docs/getting-started/intro/ (accessed 27 April 2026).

Acknowledgements

A.G.B. is supported by the Alzheimer's Association, Eisai and American Brain Foundation. P.V.H. and R.Z. are supported by NSF grant 2112533.

Author contributions

All authors meet the ICMJE requirements for authorship through (1) substantial contributions to the conception of

the work, (2) reviewing the work critically for important intellectual content, (3) final approval of the version to be published, and (4) agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Competing interests

A.M.R reports receiving research funding from GE Healthcare. A.I.L. serves as a consultant and receives fees and/or stock from EmTheraPro (co-founder), NextSense, Alamar, Cognition Therapeutics, Cognito, Eratos and Asha. The other authors declare no competing interests.

Additional information

Correspondence should be addressed to Andrew G. Breithaupt or Pedro Pinheiro-Chagas.

Peer review information *Nature Aging* thanks Vijaya Kolachalama, Laia Montoliu-Gaya, Tengfei Guo and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

[1]Goizueta Brain Health Institute, Emory University, Atlanta, GA, USA. [2]Department of Radiology and Biomedical Imaging, University of California, San Francisco, San Francisco, CA, USA. [3]School of Medicine, University of California, San Francisco, San Francisco, CA, USA. [4]Bakar Computational Health Sciences Institute, University of California, San Francisco, San Francisco, CA, USA. [5]Memory and Aging Center, Department of Neurology, University of California, San Francisco, San Francisco, CA, USA. [6]Department of Neurology, Weill Institute for Neuroscience, University of California, San Francisco, San Francisco, CA, USA. [7]NSF AI Institute for Advances in Optimization (AI4OPT), Georgia Institute of Technology, Atlanta, GA, USA. [8]H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, USA. [9]Department of Neuroscience, University of California, Berkeley, Berkeley, CA, USA. [10]Division of Biostatistics, University of California, Berkeley, Berkeley, CA, USA. [11]Center for Intelligent Imaging (ci2), Department of Radiology and Biomedical Imaging, University of California, San Francisco, San Francisco, CA, USA. [12]Department of Psychiatry and Behavioral Sciences, University of California, San Francisco, San Francisco, CA, USA. [13]Present address: H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, USA. ✉e-mail: andrew.g.breithaupt@emory.edu; pedro.pinheirochagas@ucsf.edu