

Complete Manuscript Title:

Humane Intelligence in Geropsychiatric Care: Relational Artificial Intelligence, Clinical
Wisdom, and the Moral Grid Operational Index

Short Running Title: Humane Intelligence for Geropsychiatric Care

Note: This clinical ethics and governance framework shares its name with the U.S.
nonprofit Humane Intelligence but is not affiliated with that organization.

Authors:

Helen H. Kyomen, MD, MS, on behalf of the Group for the Advancement of Psychiatry
(GAP) Committee on Aging

1. Boston University Chobanian and Avedisian School of Medicine, Boston, MA,
USA
2. Tufts University School of Medicine, Boston, MA, USA
3. Harvard Medical School, Boston, MA, USA

Corresponding author:

Helen H. Kyomen, MD, MS

• hhkyomen@bu.edu

Author Contributions (CRediT taxonomy)

- Conceptualization: Helen H. Kyomen
- Methodology/Framework design (Humane Intelligence; Moral Grid Operational Index): Helen H. Kyomen
- Investigation/Resources: Helen H. Kyomen; GAP Committee on Aging (group input)
- Writing – Original Draft: Helen H. Kyomen
- Writing – Review and Editing: Helen H. Kyomen; GAP Committee on Aging (group review)
- Visualization/Tables/Boxes: Helen H. Kyomen
- Supervision/Project administration: Helen H. Kyomen

Acknowledgments

The author thanks Kenneth M. Sakauye, MD, Iqbal “Ike” Ahmed, MD, John L. Beyer, MD, Robert P. Roca, MD, MPH, MBA, Jacob M. Appel, MD, JD, MPH, and Matthew Erlich, MD for thoughtful review and discussion that informed revisions, and the members and guests of the GAP Committee on Aging for guidance and encouragement.

Accepted Author Manuscript (AAM): This version reflects peer review and author revisions and has not been copyedited or typeset by the publisher. It is not the final Version of Record.

© 2026 American Association for Geriatric Psychiatry. Licensed under CC BY-NC-ND 4.0: <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Abstract

Artificial intelligence (AI) already influences how older adults are identified for services, supported between provider visits, and referred for care, yet most AI governance currently focuses on algorithms and infrastructure rather than the actual experiences of older adults and their caregivers. Humane Intelligence is a patient-centered, ethically attuned, relational framework for designing, evaluating, and monitoring AI in older adult care. It rests on four pillars: Relational Intelligence, Transparency with Care, Reciprocity and Consent, and Ethical Governance in Strategic Regions, and applies them from point-of-care encounters to system-level decisions.

For each pillar, guidelines follow the same sequence: signal a problem, take action, and verify benefit or harm. The framework prioritizes outcomes that matter in geriatric psychiatry, including function, distress, caregiver burden, avoidable utilization, equity, and documented harms or overrides. To protect patients, it draws a firm boundary against fully automated clinical actions and recommends clinical-grade standards for patient-facing programs, including scope-of-action labels, human-in-the-loop safeguards for high-risk situations, postmarket monitoring, and periodic certification.

This blueprint aligns with 2025 JAMA Summit on AI priorities, World Health Organization guidance for large multimodal models, United States Food and Drug Administration recommendations for predetermined change control plans, and Office of the National Coordinator for Health Information Technology decision-support intervention frameworks. The goal is practical: to translate ethics into testable patient-centered routines that clinicians can trust, healthcare systems can implement, and leaders can procure, so that AI augments rather than displaces care.

Keywords: Humane Intelligence; geriatric psychiatry; relational governance; consent; transparency; Moral Grid Operational Index; aging care.

Section 1: Introduction: Artificial Intelligence at the Threshold of Care

This article takes a hopeful stance. The aim is not to control care with rules, but to enable compassionate, accountable practice that older adults can feel.

Case Vignette

Mrs. D, 80, has just lost her husband of 60 years. Her eyes are downcast; she sits very still. Her son is beside her, unsure what to say. Hoping to help, he opens an AI chatbot and asks for options. The model responds, but the situation calls for more than the screen can hold: pacing that matches the emotional tempo of each moment, disclosure handled with care, and a human presence that can interpret the environment, acknowledge the impact of the loss, and support Mrs. D at a speed she can bear.

Artificial Intelligence is already being deployed in aging care, often without a shared, clinically grounded way to judge whether these tools are helping or quietly harming. Existing regulations and technical standards are necessary but do not tell patients, caregivers, clinicians, or leaders how to decide which AI tools are trustworthy, and when to use or decline them in the day-to-day care of older adults. Humane Intelligence provides guidelines for the design, use, and monitoring of AI so that care remains person-centered, context-aware, and accountable. It aims to optimize clinical benefit while protecting dignity, safety, and fairness.

Section 2: Defining Humane Intelligence

Humane Intelligence is grounded in both clinical ethics and the lived experience of older adults. It is a patient-centered, ethically attuned, relational framework for designing, evaluating, and monitoring AI in older adult care. It centers on what older adults actually experience, rather than on model or system-level metrics and efficiency. It also asks healthcare systems to measure what is rarely tracked: relational responsiveness, emotional appropriateness, harms avoided, and equity impacts.[1] Humane Intelligence begins with the premise that presence, attunement, clinical wisdom and compassionate moral responsibility are essential components of good care and must be built into the design and oversight of intelligent systems.[2]

What is meant by “Humane:”

“Humane” is defined here as attuned care in action. Attuned care means adjusting pace, language, and shared decisions to a patient’s cues, preferences, culture and goals, and explicitly acknowledging uncertainty, especially when safety is a concern. Attuned care in action includes the consistent expression of prosocial intent, sound social and pragmatic decision-making, emotional steadiness, reflective self-monitoring, tolerance for differing values, and explicit acknowledgment and management of uncertainty. These expressions occur within a bounded scope-of-action (i.e., clearly defined roles, responsibilities, and limits on what a person or system is expected and permitted to do). These dimensions (e.g., prosociality, social and pragmatic judgment, emotional homeostasis, reflection, value tolerance, uncertainty and ambiguity) are well-described in the wisdom literature[1] and mirror what seasoned clinicians recognize as essential to

wise, consistent care. In aging care, these are directly related to everyday needs: kindness without overstepping, practical assistance, calm demeanor, regular self-assessments, respect for family and culture, and transparent communication about what is -- and isn't -- known.

What is meant by "Intelligence:"

In this context, "intelligence" is defined as clinically informed competence within defined parameters with clear guardrails. The system's capabilities are applied to individualized clinical goals, including safety, equity, and patient preference, rather than generic benchmarks. Uncertainty is identified and documented, and the system states its confidence level with a brief rationale so that clinicians can accept, verify or defer the AI's outputs. The AI's scope-of-action is clearly delineated; for example, the system may draft clinical notes and suggest recommendations but does not place or execute orders.

The Four Pillars of Humane Intelligence with Options for Action and Indications of Successful Application:

Four pillars form the structural and ethical foundation of Humane Intelligence:

- Relational Intelligence: Attunement to the person's history, memory, emotions such as grief, culture, sensory/cognitive changes and pace; it designs for therapeutic resonance, not just task completion. Relational Intelligence recognizes that each interaction occurs within a distinct perceptual world shaped by biology, culture, and circumstance. This idea echoes the early phenomenological biology of Jakob von Uexküll, who described how every perceiving being lives within its own Umwelt -- a

surrounding world of meanings that defines what can be sensed, known, and cared for[3]. Humane Intelligence extends this recognition to AI design, emphasizing that technological perception must remain accountable to the human worlds it touches. In practice, this means matching interventions to each patient's Umwelt, rather than assuming that one interface or script fits all. (See Vignette C for case → problem → Humane Intelligence solution.)

- **Transparency with Care:** Truthfulness delivered in plain language with humility; it encompasses multiple stratified explanations, uncertainty, and parameters that older adults and caregivers can utilize with greater ease.[4] (See Vignette A for how “performative empathy” without uncertainty disclosure can violate beneficence at the point-of-care and for a description of a Humane Intelligence solution.)
- **Reciprocity and Consent:** Consent is an ongoing process with choices regarding tasks and data-use, regular updates, and straightforward opt-out options, which can be recorded as consent receipts within electronic health records (EHRs).[5] (Vignette A illustrates layered, revocable consent and escalation for high-risk queries.)
- **Ethical Governance in Strategic Regions:** Strategic regions are high-risk, high-leverage parts of the healthcare system where AI-enabled decisions concentrate both authority and risk -- for example, geriatric psychiatry units, long-term care facilities, and payer management systems -- so that seemingly small changes can have large effects on older adults' safety, access, and equity status. In such strategic regions, safeguards are matched to risk level; explanations are plain and transparent, making the design, function and decision-making processes of healthcare technologies clear and understandable. Independent, structured

evaluation occurs before and after AI deployment. Those who are most affected review AI technologies and policies, and audits are deeper where power imbalances are greatest due to authority, knowledge, and resources being unevenly distributed (such as in inpatient geriatric psychiatry, long-term care, and payer-driven service management).[6] (See Table 2 for monitoring; Vignette B demonstrates the consent stakes when systems-level tools shape discharge decisions.)

Core Values Anchoring the Framework

In addition to its four pillars, Humane Intelligence is grounded in core clinical and moral values that echo through geriatric psychiatry and caregiving:

- Presence over performance: Orient to the person, then the task.
- Recognition over prediction: Validate the concern, then model the risk.
- Compassion over control: Meet the patient where they are, then work together collaboratively.
- Dignity over dominance: Support preference and pace, over expediency, convenience or profitability.
- Mutuality over extraction: When the system uses a person's data, the person receives a clear benefit, such as clearer, plain language explanations or information, tailored support or better care. Data use benefits should not be one-way.

These values reflect what is often most healing in psychiatric and elder care: the human experience of being seen, known, and treated with respectful care. Artificial Intelligence systems, if they are to participate in this space, must be designed with these imperatives in mind.[7,8]

Two areas of AI use in care are distinguished: (1) AI for systems -- tools that shape workflows, financing, triage, and population management; and (2) AI for individuals -- companions and decision aids that meet a person at the point of care. The Humane Intelligence Four Pillars span both places of use. In practice, Humane Intelligence is less about creating a single “best” AI system and more about wisely selecting and governing tools, so that each older adult receives the tool that best fits their world, addresses real needs, and minimizes unintended harms.

To move from concerns to action, a shared map can be helpful. The Moral Grid Operational Index can serve as that map – clarifying risk and benefit not only in how AI tools work, but in how they affect relationships, consent, and change.

Conceptual Language for Ethical Artificial Intelligence Development: The Moral Grid Operational Index

The Moral Grid Operational Index (see Table 1) links components of each Humane Intelligence pillar to specific point-of-care behaviors and to concrete, auditable evidence. It draws on the National Institute of Standards and Technology Artificial Intelligence Risk Management Framework 1.0 (NIST AI RMF 1.0)[9], and the NIST AI RMF Playbook,[10] translating their “Govern, Map, Measure, and Manage” functions into a healthcare-relevant profile of observable practices.

(Insert Table 1 here.)

The Moral Grid Operational Index makes the four pillars of Humane Intelligence -- Relational Intelligence, Transparency with Care, Reciprocity and Consent, and Ethical Governance in Strategic Regions -- operational. It specifies day-to-day behaviors and the kinds of evidence that show that those behaviors occurred. The grid supports practical decisions about safety, respectful engagement with older adults, transparent clinical tools, renewal of consent for cognitively vulnerable patients, attention to social determinants of health, and accountable governance across connected systems. In this way, abstract ethics become commitments in clinical care and AI design across different perceptual worlds or *Umwelten*[3].

Table 1 focuses on point-of-care behaviors. Building on that foundation, Table 2 introduces a risk-tiered view, showing how safeguards escalate across risk tiers T1-T3. Operational indicators include uncertainty alerts, consent renewal, changes in equity monitoring, audit depth, and ethics review committee consultation when consent is unclear or contested. At any tier, an equity indicator may signal patterns of unequal benefit or burden.

(Insert Table 2 here.)

The Moral Grid terms and concepts are meant not only to guide design conversations, but also to provide shared language that clinicians, ethicists, caregivers and leaders can use to evaluate emerging systems that claim to offer care.

Seniors Uniting Nationwide to Support Health, INtegrated Care, and Evolution (SUNSHINE) and Humane Intelligence -- How They Fit Together

The SUNSHINE framework[11] describes how communities and healthcare systems can use AI to strengthen resilience in aging care at scale. Humane Intelligence sits closer to the point-of-care: it specifies what older adults and clinicians actually experience and the metrics and guardrails that keep those experiences safe, clear, reviewable, and improvable. The Moral Grid Operational Index links SUNSHINE and Humane Intelligence by making both system-level intentions and point-of-care practices visible and auditable.

(Insert Table 3 here.)

Consensus from the 2025 JAMA Summit on AI converges on four system-level priorities for healthcare AI: (1) multistakeholder lifecycle management (where diverse interested parties such as clinicians, patients, caregivers, data scientists, quality-improvement specialists, and leadership work together at each stage of problem definition, training and validation, deployment, post-deployment monitoring, and retirement), (2) evaluation and monitoring tools that measure effectiveness, not only safety, (3) a nationally representative learning data infrastructure, and (4) aligned incentives to improve feasibility.[12] As detailed in Box 1, these priorities harmonize with the Humane Intelligence pillars and provide operational scaffolding for bringing Humane Intelligence concepts into routine care.

(Insert Box 1 here.)

Section 3: Clinical Relevance in Geriatric Psychiatry

Vulnerabilities of the Aging Population and Ethical Responsibility

Older adults often face intersecting vulnerabilities that heighten ethical risk:

- Cognitive changes (e.g., mild or major neurocognitive disorders) may impair informed consent, memory of prior disclosures, or recognition that one is interacting with AI.[13]
- Sensory impairments (e.g., visual, auditory or tactile loss) may reduce access to, or distort perception of, interfaces not designed for older adults.[14]
- Psychosocial risks such as loneliness, bereavement, isolation, depression or shame may increase susceptibility to emotionally misleading or nontransparent interactions.[15]
- Institutional or relational power differentials may inhibit older adults from voicing concerns, asking questions, or refusing AI-based recommendations in clinical or caregiving settings.[16,17]

These vulnerabilities amplify the ethical responsibility to ensure that AI systems not only perform accurately, but also attune, communicate clearly, and honor the personhood of the patient. Thus, the relevance of Humane Intelligence to geriatric psychiatry is immediate and growing. As AI tools are increasingly deployed across health care systems, older adults are often among the first populations targeted for implementation

due to caregiver shortages, cost containment pressures, and the perception that they may benefit from passive monitoring, behavioral modeling, or digital companionship. Yet this population also is uniquely vulnerable to the unintended consequences of poorly attuned, opaque, or extractive systems.[18,19]

Examples of Current and Emerging Artificial Intelligence Use in Older Adults

Artificial Intelligence technologies already present in the care of older adults include:

- Companionship/virtual attendants designed to offer companionship or therapeutic conversation (e.g., ElliQ, in pilot and implementation studies, was associated with reduced loneliness and high engagement).[20,21]
- Fall detection/behavioral analysis in memory care (e.g., computer-vision fall detection was associated with faster response times; agitation detection technologies may help prevent falls in the cognitively impaired).[22,23]
- Automated medication adherence (e.g., smart pill dispensers, voice assistants for reminders and after visit debriefs).[24,25,26]
- Remote patient monitoring (e.g., wearables, speech analysis, video monitoring in hospitals and long-term care for health monitoring, may lead to more timely intervention and improve disease management).[22,23]
- Predictive modeling tools to assess risks, such as hospitalization, emergency department use, or cognitive decline[27]
- Automated clinical documentation tools that generate progress notes or assessments using ambient audio (e.g., Nuance DAX, DeepScribe)[28,29]

While some of these tools may offer potential benefits -- including expanded access, increased efficiency, and reduced staff burden -- they also raise critical concerns around oversight, relational disruption, and equity. For instance, an AI system that flags a patient as “nonadherent” may do so without incorporating contextual information such as grief, sensory loss, or treatment side effect burden. A chatbot offering emotionally responsive dialogue may simulate empathy without disclosure or the capacity for correction -- creating ethical ambiguity about the boundaries of care.[30,31] In all such cases, the relational, emotional, and cognitive dimensions of aging must be accounted for in the system's design, deployment, and clinical integration.

Psychotherapeutic Models as Guides

The field of geriatric psychiatry offers robust traditions that can directly inform the development of Humane Intelligence:

- Psychodynamic models emphasize attunement, affective resonance, transference dynamics, and unconscious communication -- all of which may shape how older adults respond to emotionally expressive or simulated systems.[32]
- Cognitive-behavioral frameworks address the role of belief structures, cognitive distortions, and interpretation of feedback -- especially relevant when considering AI-generated suggestions, prompts or diagnostic predictions.[33]
- Narrative and meaning-centered therapies emphasize the importance of identity, continuity and life review -- values that can inform memory-sensitive and person-centered AI interactions.[34,35]

- Creative arts, spiritual, and psychosocial therapies highlight cultural context, symbolic meaning, and aesthetic resonance -- offering critical input on AI interface design and emotional resonance.[35]

These models demonstrate that care is not merely transactional, and that therapeutic relationships -- even those mediated or augmented by AI -- require moral creativity and relational skill. Humane Intelligence offers a structured framework to bridge the technical and therapeutic domains of AI use in geriatric psychiatry. It provides a vocabulary and ethical orientation for clinicians, developers, and policy stakeholders to evaluate AI systems not only by what they do, but by how they relate -- and whether they contribute to healing, autonomy, and dignity of the person.

Section 4: Composite Case Vignettes

These cases demonstrate how the Humane Intelligence pillars and Moral Grid principles may operate at the point-of-care and during service decisions. These illustrative composite cases are synthesized from clinical patterns and published reports; they do not describe specific patients. A brief vignette opened the article to set the tone; here, fuller cases are presented now that the Humane Intelligence framework has been laid out.

Vignette A -- Individual Level (Relational Intelligence, Transparency with Care, Reciprocity and Consent)

Mrs. T and the “helpful” reassurance: An example of well-intentioned alignment that increases risk, and how Humane Intelligence redirects the system.

- Before Humane Intelligence Principles Are Implemented:

Mrs. T, 78, lives alone with mild cognitive impairment and recent grief after her sister’s death. Nights are the hardest. Her clinician recommends an AI chat companion embedded in the patient portal to help with routines and sleep hygiene between visits.

At 2 a.m., Mrs. T asks if it’s okay to “take one extra” of her sedative because the first dose did not help. The model, tuned to be soothing, offers confident reassurance without checking the prescription details or age-related risks. Mrs. T doubles the dose, awakens groggy, falls on the way to the bathroom, and arrives at the clinic shaken and bruised.

- After Humane Intelligence Principles Are Implemented:

At follow-up, the team recognizes a gap: the AI model “performed empathy” but violated beneficence and honesty by offering comfort that implied safety. The clinic reconfigures the system guided by Humane Intelligence:

- Scope and routing: Medication dose advice is out of scope of the AI. These queries are blocked and routed to an on-call clinician.
- Uncertainty and redirection: Policies require explicit uncertainty and redirection; the AI models must disclose that it cannot answer and guide Mrs. T to human help.
- Consent and follow-up: The AI’s capabilities and limits are explained; informed consent is obtained, and nightly check-ins encourage appropriate escalation rather than only medication dose changes.

- How to know it worked: Reduced medication-related escalation; decreased after-hours fall events; improved patient trust and understanding scores; and audits showing that high-risk queries are blocked or escalated rather than answered.

Vignette B -- Systems Level (Transparency with Care, Ethical Governance in Strategic Regions)

The discharge score and the quiet exclusion: An example of when incomplete information and missing context lead to inappropriate decisions, and how remediation is initiated.

- Before Humane Intelligence Principles Are Implemented:

A large hospital deploys an AI discharge-readiness score to predict who can go home safely with limited services. Length of stay drops and executives are pleased.

Months later, a geriatric psychiatry liaison notices a pattern: patients with hearing and vision impairments -- disproportionately older adults -- are scored “ready for discharge” sooner based on proxies such as oriented conversation length (minutes a patient sustains coherent, oriented dialogue) and call-light frequency (nurse call button presses per shift). Because hearing and vision sensory loss lower measured conversation length and call-light use, the model mislabels readiness. Readmissions and emergency department (ED) returns increase in this subgroup, but the overall length-of-stay averages still look acceptable.

- After Humane Intelligence Principles Are Implemented:

A governance review finds no pre-deployment bias assessment for sensory impairment, no independent evaluation records and logs, and minimal postmarket monitoring by subgroup. The hospital initiates a governance plan for this strategic region:

- Independent evaluation: An independent body reviews the model and publishes evaluation records and logs, including plain-language documents describing limits.
- Subgroup monitoring and rollback: Outcomes are monitored by age, sensory impairment, language, and housing, with triggers to roll back the model or workflow when metrics worsen.
- Remediation: Discharges in flagged groups require human review before finalization.
- Portable auditing: Preference is given to open protocols that support portable auditing.
- Certification: A certification mark is contingent on quarterly bias checks.
- How to know it worked: Lower readmission rates in affected subgroups; improved test characteristics by impairment status; decreased number of and faster closure of bias incidents; and an appropriate rate and timeliness of clinician overrides for flagged cases.

Vignette C: A Story of Attuned Artificial Intelligence Assisted Care

This composite case shows how the same type of AI can either fragment or facilitate attuned care, depending on design and deployment.

- Before Humane Intelligence Principles Are Implemented:

Mrs. D is an 82-year-old woman living in an assisted living facility. She lost her husband two years before moving in. She has a long history of depression and anxiety, now compounded by grief, hardness of hearing, low vision and mild cognitive impairment. Staff describe her as “pleasant but withdrawn.” She often sits alone in her room with the television on but not really following it. She does not engage in available scheduled activities. She is adherent with offered medications but misses meals, and her weight is dropping.

As part of a new program, the facility installs a virtual assistant in each resident’s room -- an AI-enabled conversational agent designed to provide daily prompts, companionship, and reminders. There are illustrative examples of similar products in elder care with performance varying by implementation. For example, New York State’s Office for the Aging piloted the ElliQ companion with resultant associations with reductions in loneliness; other pilots are exploring proactive conversational supports at home.[20,21]

When first activated, the device greets Mrs. D with a cheerful voice: “Good morning! It’s time to start the day. Don’t forget to eat a healthy breakfast!” Mrs. D does not respond. The virtual assistant repeats the prompt twice over the next 30 minutes. Later, it offers trivia questions and a weather report. By the end of the week, the staff notes that Mrs. D remains withdrawn, has not engaged with the virtual assistant, and appears more fatigued.

- After Humane Intelligence Principles Are Implemented:

A psychiatrist is consulted. After meeting with Mrs. D and reviewing her history, the psychiatrist collaborates with the care team to adjust the virtual assistant's settings and integrate Humane Intelligence principles into its use. The virtual assistant is reprogrammed to recognize long pauses, respond to lower volume speech, and deliver prompts in simpler language at a slower pace with a warmer tone. Instead of giving directives, the virtual assistant now offers gentle choices: “Hi, Mrs. D. I’m here if you’d like to talk. Would you like to hear a story, listen to some music, or just sit quietly with me?”

The virtual assistant also is linked to a personalized database of Mrs. D’s life story -- family members, favorite music, and past meaningful places. When Mrs. D eventually responds with, “I don’t know,” the assistant gently says: “It’s okay not to know. I’m glad to be with you. Today, in Rockport, it is sunny with a breeze; that’s where you used to go with Mr. D, right?” This creates a bridge between present prompts and remembered experiences.

Within three weeks, teach-back comprehension and mealtime attendance improve. In case review, the care-team credits the changes in tone, pacing and consent renewal prompts as helpful. The assistant does not replace therapy or human care, but it becomes a bridge, a companion presence shaped by clinical insight and ethical design that augments, rather than displaces, human support.

Interpretation and Ethical Insight

This vignette illustrates the difference between a generic AI system and one shaped by Humane Intelligence. In the first part, the virtual assistant operates according to standardized routines, oblivious to emotional nuance, history, or context. In the second,

clinical input reshapes tone, rhythm, and responsiveness, allowing the system to meet Mrs. D as a person navigating grief, aging, and loss. By embedding relational attunement, Transparency with Care, and therapeutic reciprocity, the AI assistant moves from transactional reminders to attuned accompaniment.

These vignettes do not suggest that AI can or should replace human contact. Rather, they highlight how ethical framing and clinical insight can reshape AI systems -- moving from simple tools with transactional interactions, to relationally attuned supports, to genuine augmented care.

Section 5: Ethical Governance in Strategic Regions in Practice: From Policy to Point-of-Care Reliability

Building on the vignettes, this section shows how the four pillars of Humane Intelligence can be translated into concrete, patient centered ways of evaluating specific AI tools. Governance, in this context, is care in action: proportionate, relational, and designed to widen safe capability rather than simply restrict it. Sections 1–4 described challenges in incorporating new technologies into the mental health care of older adults and demonstrated how Humane Intelligence can shape specific clinical situations. Sections 5-7 extend that foundation into systemic practice: how to design, procure, and govern systems so they behave in responsible, standards-aligned ways in clinical settings, especially in aging care.

A companion research concept, Relational Evaluation of AI in Clinical Horizons – Long Conversation (REACH-LC), focuses on AI used in extended, multi-session dialogues. It adapts Learning Health System principles[36] to AI assurance across the lifecycle, with particular attention to long-conversation safety. (See Appendix A.)

5.1. Ethical Governance in Strategic Regions

High-risk areas such as geriatric psychiatry units, long-term care, and payer management systems concentrate both risk and authority. Without adequate oversight, AI can magnify inequities. Humane Intelligence recommends shifting governance from compliance-only models to ones that emphasize relational safety (the felt sense of being known, respected, and safe within the care relationship), supported by systems that are transparent, reciprocal, and auditable.

At the policy level, the Moral Grid Operational Index implements widely endorsed governance norms -- risk-based oversight, transparency documentation, consent traceability, and continuous audit -- so healthcare systems can translate abstract AI rules into point-of-care safeguards for older adults. The Food and Drug Administration's (FDA's) Predetermined Change Control Plan (PCCP) framework offers an emerging baseline for lifecycle oversight of AI-enabled medical devices and may influence how healthcare systems and regulators evaluate updates over time.[37] The scope-of-action for any AI tool should be defined up front, and updates should be governed through a PCCP-style change control (a structured process to manage changes so that a product

stays safe and compliant) and be communicated through versioned evaluation records and logs.[37]

Business-operations AI also can reshape access and quality. Tools for scheduling, bed management, prior authorization, and revenue cycle management may influence who receives care, how quickly and under what conditions.[9] Even when these tools are not regulated as medical devices, Humane Intelligence treats them as clinically relevant. At minimum, health systems should track their impact on patient-important outcomes by age and other equity relevant groups, and be prepared to pause or rollback tools if outcomes worsen.

Clinical-Grade Use: Testing and Certifying/Recertifying/Licensing Counseling-Oriented AI

Because AI increasingly mediates therapeutic moments, counseling-oriented AI programs should meet a therapeutic standard comparable to other clinical tools.

Certification and periodic recertification -- tied to scope-of-action labels (i.e., brief, visible statements that specify the AI tool's allowed actions and limits, and whether human oversight is needed), human-in-the-loop requirements for higher-risk use, and postmarket monitoring -- are recommended to align AI with medical-device and digital-therapeutic norms, adapted to behavioral health and aging care.[6,27,38,39] This approach safeguards innovation while protecting patients.

Ethical Oversight

When AI is deployed under conditions of uncertainty, evaluation may require ethics review committee consultation in addition to quality-improvement review. If evidence accumulates that teams repeatedly fail to use a proven tool or to follow guardrails, then standard-of-care concerns may be raised. Humane Intelligence thus links Reciprocity and Consent to clear accountability both for decisions to use an AI tool and decisions not to use it when it is available and indicated.

Emerging programs, such as The Joint Commission–Coalition for Health AI (TJC-CHAI) collaboration, are beginning to define voluntary AI certification pathways for hospitals and other facilities across the continuum of care.[40] In parallel, national policy can align incentives and support a federated, collaborative learning network with geriatric sentinel sites to ensure generalizable, equity-positive results[12]. Oversight follows what is measured: function, distress, caregiver burden, avoidable utilization, equity improvements, and documented harms or overrides, each with a trigger for review and a clearly named accountability node.

(Insert Table 4 here.)

Post-deployment Labs

The National Evaluation System for Health Technology (NEST) represents a maturing national infrastructure for real-world evidence about medical devices and provides a model for lifecycle oversight of clinical AI tools.[41] Participation in national AI evaluation networks (“post-deployment labs” or shared testing collaboratives) is recommended to

share methods, benchmarks, and post-deployment findings within a learning-health-system model.[36,41] For aging and mental health, these networks can function as laboratories for Humane Intelligence in practice, testing how specific guardrails, consent processes and subgroup monitoring strategies perform across diverse sites and re-introducing improvements back into local systems.

5.2. Ethical Governance in Strategic Regions and Patient Safety

Good governance includes the clinical practice of making technology safe for patients. Before any AI system is introduced, a pre-deployment review should be conducted, including with input from individuals with relevant lived experience. Governance boards should include older adults, caregivers, frontline clinicians, technical reviewers, bioethics and safety or quality representatives, and legal assessors. Their perspectives help set acceptable risk and usability thresholds.[38,39]

The pre-deployment review should confirm three basics. First, usability: the AI tool must be accessible and learnable, with minimal steps and an interface that supports clinicians under time pressure. Second, data provenance and model lineage: hospitals should know where the data originated, how it was processed, and how the model has changed over time. Such information also supports privacy and security oversight. Third, capability limitations: as with medications, what the AI system cannot do, and where its outputs should not guide care, should be clearly stated. Performance and fairness thresholds should be met in the populations served, including older adults and

caregivers, with specific attention to frailty, cognitive impairment, language barriers, and cultural needs.[42,43]

Together, this pre-deployment review confirms that the AI tool fits local workflows, that data and model lineage are documented, and that intended uses and clear limitations are visible to clinicians and patients. Box 2's Criteria– Accountability–Support–Recourse (CASR) checklist offers a concise way to embed these requirements into procurement and quality-improvement processes, and to link them with payment arrangements so that safe improvement is rewarded.[38,39,42,43]

(Insert Box 2 here.)

After pre-deployment review, AI systems can be introduced through a staged rollout with continuous monitoring for performance, bias, and safety. Section 7 describes a three-stage roadmap -- Seed and Stabilize, Align and Embed, and Improve and Share -- that services can adapt to their local context.

Section 5.3. Clinical Safety: From Taxonomy to Triage Protocol

Teams, leaders and vendors all have roles in making AI-supported care safe. Guardrails matter most where automatic responses fail. A clear boundary keeps final clinical actions in human hands: AI may inform, summarize, and forecast, but must not order, prescribe, discharge, detain, or deny care without a responsible clinician's verification and signature.[37]

Rather than relying on broad metaphors such as “hallucinations” or “sycophancy,” teams can label specific failure types -- fabricated content, misattribution, omission, decontextualization, overreach, calibration error, and agreeableness bias – and attach actionable next steps. These include highlighting uncertainty, routing high-risk cases to humans, and limiting scope-of-action to approved guidelines. Structured tagging and response plans may help prevent error cascades and maintain trust. Table 5 summarizes these failure tags and documentation cues for incident logs.

(Insert Table 5 here.)

Section 6: Reducing the Risk for Harm and Promoting Safety

Effectiveness, Not Only Safety

Beyond standard safety and process checks, Humane Intelligence argues for and recommends prospective evidence that use of an AI tool in older adults across settings improves patient-important outcomes (e.g., function, distress, caregiver burden, avoidable utilization, changes in equity).[12] Therefore, safety monitoring is paired with embedded clinical trials (e.g., with stepped-wedge, platform or quasi-experimental study designs) combined with ongoing local validation.

AI is likely to disrupt every part of health care delivery in the coming years.[12] It is already embedded in after-hours triage, electronic health record (EHR) documentation,

decision support procedures, home and community care monitoring, and long-term care.[27] Properly designed, AI can extend telepsychiatry to rural or homebound older adults, provide hearing- and language-friendly summaries of discharge instructions, automate refills and prior authorization packets, and draw attention to clinically meaningful changes (e.g., fluctuations in vital signs, sleep or appetite)[9] so that health care practitioners can focus their efforts to where they are needed most. Where care is relational, however, safety must include attunement to the person, consent-as-process, and the capacity for correction.

As an adjunct, not a replacement, AI can offer ethically framed companionship through context-aware check-ins, personalized music or life-story prompts, and the gentle provision of daily structure.[20,21] However, these same features can be harmful if bias, dehumanization, opacity, and power asymmetries go unaddressed. AI models trained primarily on younger, English-speaking populations may under-detect depression in older adults and other underrepresented groups.[44] Grief, sensory loss, or cognitive impairment may hinder older adults from responding to queries appropriately, resulting in outreach that may be perceived by some as punitive. Vendor non-disclosure agreements (NDAs) or black-box designs (where a system's internal workings are not apparent or considered "opaque") may limit the ability to challenge errors.[27,45] Sensory or cognitive changes or frailty can lead to motoric, response speed and other performance related difficulties in accessing and utilizing AI tools correctly.[18] Older adults' values, preferences, and personhood may be quietly subordinated to others'

convenience, or to institutional and other systemic priorities.[19] These dynamics produce tensions that geropsychiatric practice must address.

These tensions may be alleviated by promoting truthfulness, guarding against exploitation, and favoring a non-directive stance.

- Truthfulness can be promoted in many ways, such as linking claims to their sources, flagging uncertainty with “uncertainty badges” or “uncertainty indicators,” avoiding confident-sounding fabrications, and discouraging overly aggressive auto-correction prompts.
- Exploitation can be prevented by using scam/fraud detection prompts, providing links to trusted resources, and routing to paid services only after a person-initiated request and double confirmation.
- A non-directive stance is inherent when Person-Centered Therapy[46] or Motivational Interviewing[47] techniques are used, and when giving advice requires crossing a documented certainty/consent threshold and includes alternative choices with the benefits and risks of each option.

Guardrails for Aging Care – Adapted from The World Health Organization’s (WHO’s) Large Multi-Modal Models Guidance[2]

To surpass the theoretical and translate principle into practice for a vulnerable population, guardrails proportionate to clinical risk can be specified, observed, and embedded as part of routine care:

1. Scope-of-Action Labels -- Every program declares what it can and cannot do (assistive, advisory, guided co-production; no fully automated clinical actions).
2. Hard No-Advice Zones -- Dose changes, crisis guidance, involuntary care, and eligibility determinations route immediately to a responsible human clinician.
3. Uncertainty Disclosure -- Visible indicators and plain-language rationales indicate when confidence is low or evidence is contested; teach-back is used to confirm comprehension.
4. Consent-as-Process -- Layered, revocable consent with periodic updates is documented; consent receipts are recorded in the EHR and viewable in the patient portal.
5. Subgroup Monitoring with Rollback -- Outcomes are stratified by age bands, sensory/cognitive impairment, language, disability, and housing; pre-set triggers roll back the model or workflow when equity metrics worsen.
6. Versioned Transparency Artifacts -- Public evaluation records and logs, model/dataset lineage (the sources, curation and versions of the AI model and dataset over time), and change-control notes (what changed, why, expected impact, validation results, and new risks and mitigations for each update that could affect care) are documented.
7. Independent Oversight and Redress (i.e., repair) -- Third-party or cross-site audit, incident reporting, and a standing recourse pathway (Stop/Undo/Explain/Redress) with service-level timelines are implemented.
8. Verification Hooks: Falls after medication prompts, blocked-vs-answered high-risk queries, 30-/90-day ED revisits, caregiver burden, teach-back pass

rates, and changes in equity are reviewed on a fixed cadence with accountability nodes named in Table 4.

It is important to note that general purpose chatbots are not crisis tools; when risk rises, safeguards can degrade.[48] They are not designed, tested, or approved to function as crisis tools, and should not be relied upon as primary interventions in situations of imminent or escalating risk to self or others (see Box 3, consistent with WHO guidance on safeguards for high-risk use).[2] Licensed clinical applications with human escalation should be used instead. Institutional policies should reflect these patient-safety requirements.

(Insert Box 3 here.)

Section 7: A Roadmap to Implementation

Section 7 extends governance into daily operations: Straightforward, two-way explanations at the point-of-care, risk-proportionate safeguards, and a change-control system that treats updates like clinical events which are logged, reviewed, communicated, and ready for correction. The following is a sample guide to enactment. Time frames for implementation will vary by site size, integration complexity, and risk tier.

Stage 1: Seed and Stabilize[9,12,36]

- Identify a specific workflow to pilot (e.g., caregiver education after ED discharge).
- Establish consent receipts and uncertainty indicators that are logged with confidence levels and brief rationales, and train staff in their use.
- Collect baseline data on comprehension, adherence, and return-visit rates.

Stage 2: Align and Embed[9,12,27]

- Incorporate feedback loops that update prompts, policies and guardrails.
- Initiate an independent audit and release a plain-language summary.
- Broaden scope-of-action to an additional workflow (e.g., social isolation screening and referrals).

Stage 3: Improve and Share[9,12,27]

- Integrate Criteria, Accountability, Support, and Recourse (CASR) procurement levers into all vendor agreements.
- Present outcome reports to the governance board and stakeholder community.
- Develop plans for replication across sites.

Deliverables: For each stage, core deliverables include a one-page practice brief, a transparent change log, and patient/caregiver handouts.[9,42]

(Insert Box 4 here.)

7.1. Evaluation and Outcomes: Measuring What Matters

Outcomes can be measured in several complementary ways. For example:

- Process Measures: consent completion and comprehension (teach-back pass rate), percentage of encounters displaying an uncertainty indicator, escalation rates.

- Clinical Measures: adherence to safety instructions, 30- and 90-day ED revisit rates, changes in standardized scale scores such as PHQ-9[49] or GAD-7[50] for mood when applicable.
- Relational Measures: patient-reported trust and perceived understanding, caregiver burden, clinician moral distress.
- Equity Measures: stratify by age (65–74, 75–84, 85+), language, disability status, and socioeconomic status; preregister analyses; publish negative results.
- Link to the Moral Grid Operational Index: align outcomes with grid dimensions (e.g., presence-related behaviors; Transparency with Care; dignity, autonomy, and consent integrity; access, inclusion, cost transparency and equity delta monitoring; safety and supervision; data stewardship). Ensure indicators match behaviors needed for procurement and audits.

Responsible deployment requires proving improved outcomes across settings, not merely documenting model behavior. Consistent with guidance from the 2025 JAMA Summit on AI, embedded clinical trials (e.g., with stepped-wedge, platform or quasi-experimental study designs) and ongoing local validation during use are recommended so that technological performance, human-computer interaction, training, and context are evaluated together.[12]

7.2. Safeguards

Humane Intelligence offers significant potential, but risks remain, including model drift, changing contexts, and human factors. Organizations should define capability limits for high-risk processes, require human verification for irreversible decisions, and introduce

sunset clauses when equity or performance benchmarks are unmet. These safeguards should be supported by robust change-control systems and audit documentation that clinicians can understand and use. Implementing CASR in contracts, using consent receipts, surfacing uncertainty, and involving older adults and caregivers in governance may make innovation more trusted and effective.

8. A Call to Action -- Advancing Humane Intelligence in Mental Health and Aging Care

SUNSHINE is a system-level framework designed to advance population-level resilience and integration across services. Humane Intelligence sits one layer closer to the person: it helps constrain and make safe the moment-to-moment actions of AI systems that interact with older adults and families, and it clarifies what vendors must build, test, and disclose. Together, they are complementary: SUNSHINE moves the system; Humane Intelligence protects the person more directly.

The Moral Grid Operational Index demonstrates that Humane Intelligence need not remain abstract. It can be lived through relationship, transparency, consent, and governance in everyday care. The future of AI in mental health and aging care will depend on decisions made now. Humane Intelligence is offered as an ethical framework for these choices: designing for presence, governing for dignity, and measuring what matters in care.

An important step is to translate the components of Humane Intelligence into specific responsibilities for clinicians, developers and leaders. Transforming such principles into actionable roles can enhance the safety and quality of care for older adults:

For Clinicians and Care Teams:

- Inquire whether AI tools can handle emotional nuance, complexity, and individual histories.
- Advocate for relational compatibility as well as technical performance.
- Use clinical wisdom to judge when AI enhances or undermines relationship, autonomy, and personhood.
- Teach trainees to evaluate both outcomes and processes -- and their ethical alignment.
- Build clinician competencies: Construct core AI literacy into training and continuing education, including uncertainty communication, escalation criteria, consent-in-context, subgroup equity checks, and safe override practices.

For Developers and Technology Partners:

- Include clinicians, caregivers, and older adults in participatory design and evaluation.
- Prioritize transparency, correctability, and emotional attunement alongside technical performance.
- Treat relational dynamics, ethical context, and cognitive diversity as core design constraints, not afterthoughts.

For Policymakers and Institutional Leaders:

- Support interdisciplinary ethics review mechanisms for AI in aging and mental health.
- Invest in comparative research on AI with versus without explicit relational and ethical parameters.
- Align procurement with values such as dignity, informed consent, and clinical relevance -- not only cost or novelty -- using instruments such as CASR and Moral Grid-aligned metrics.
- Establish accountability structures that include those most affected, especially older adults in institutional or marginalized settings.

For All Stakeholders:

Humane Intelligence speaks not only to technological advancement, but to the values guiding care for older adults -- presence, respect, protection from harm, and recognition of personhood. It cautions against reducing care to metrics or economics alone and calls us back to attuned relationships and a professional obligation to heal. The crucial task is to create not only smarter systems, but wiser ones -- relationally attuned, transparent with care, and reciprocal in consent.

This blueprint is offered not as a fixed doctrine but as an adaptable guide, inviting participants to develop systems that honor core human values, do not forget the heart and soul, remember what matters, and integrate thoughtful, transparent approaches to

AI in aging and mental health care. The hope is that the elders who shaped the world we share are accompanied with humility, compassionate care, and vision.

Limitations and Appropriate Use

The Humane Intelligence blueprint is a practical guide to the safe and accountable AI use at the point-of-care. It is not a substitute for clinical judgment or local policy.

Although it is aligned with current standards and regulations, evidence and rules will continue to evolve; organizations should review the latest guidance and adapt implementation accordingly. The box- and table-level tools are intended to augment clinical workflows, consent, and oversight -- not to automate diagnosis, triage, or crisis management. General-purpose chatbots are not appropriate for emergencies; routing to licensed clinical services remains essential. Examples and measurement hooks illustrate feasibility across settings; but require local validation, equity review, and periodic re-evaluation as models, data, and populations change.

Footnote: Umwelt (literally “surrounding world”) is von Uexküll’s term for the unique, meaning-laden world each perceiver inhabits -- a world not merely observed but composed through relationship, attention, and need[3].

Declaration of Generative AI and AI-assisted Technologies in the Writing Process:

During the preparation of this work the author(s) used ChatGPT-5 Pro in order to assist with language editing, formatting, and organizing the manuscript. After using this tool/service, the author(s) reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

1. Meeks TW, Jeste DV: Neurobiology of wisdom: a literature overview. Arch Gen Psychiatry 2009; 66(4):355–365. doi:10.1001/archgenpsychiatry.2009.8
2. World Health Organization: Ethics and governance of AI for health: guidance on large multi-modal models (LMMs). Geneva: World Health Organization; 2024. Licence: CC BY-NC-SA 3.0 IGO.
3. von Uexküll J: A Foray into the Worlds of Animals and Humans, with a Theory of Meaning. Translated by O’Neil JD. Minneapolis, MN: Univ. of Minnesota Press, 2010 [orig. 1934]
4. Talevski J, Wong Shee A, Rasmussen B, et al: Teach-back: a systematic review of implementation and impacts. PLoS One 2020; 15(4):e0231350. doi: 10.1371/journal.pone.0231350
5. HL7: FHIR R5—Consent resource. 2023. (Official spec.)

6. European Union: Regulation (EU) 2024/1689 establishing harmonised rules on artificial intelligence (AI act). Official Journal of the EU. Published June 13, 2024. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>. Accessed November 1, 2025
7. Sinclair S, Beamer K, Hack TF, et al: Sympathy, empathy, and compassion: A grounded theory study of palliative care patients' understandings, experiences, and preferences. *Palliat Med* 2017; 31(5):437-447. doi:10.1177/0269216316663499
8. Malenfant S, Jaggi P, Hayden KA, et al: Compassion in healthcare: an updated scoping review of the literature. *BMC Palliat Care* 2022; 21(1):80. doi: 10.1186/s12904-022-00942-3
9. National Institute of Standards and Technology (NIST): Artificial intelligence risk management framework (AI RMF 1.0). Gaithersburg, MD: NIST; January 2023. (Official PDF.)
10. National Institute of Standards and Technology (NIST): NIST AI RMF Playbook. Gaithersburg, MD: NIST; 2023. https://airc.nist.gov/AI_RMF_Knowledge_Base/Playbook. Accessed November 1, 2025
11. Chen J, Maguire TK, McCoy RG, et al: Reimagining resilience in aging: leveraging AI/ML, big data analytics, and systems innovation. *Am J Geriatr Psychiatry* 2025; 33(9):1005–1017. doi:10.1016/j.jagp.2025.05.007

12. Angus DC, Khera R, Lieu T, et al: AI, health, and health care today and tomorrow: the JAMA summit report on artificial intelligence. *JAMA*. 2025; 334(18):1650–1664. doi:10.1001/jama.2025.18490
13. Appelbaum PS: Assessment of patients' competence to consent to treatment. *N Engl J Med* 2007; 357(18):1834–1840. doi:10.1056/NEJMcp074045
14. Czaja SJ, Boot WR, Charness N, et al: Designing for older adults: principles and creative human factors approaches. 3rd ed. Boca Raton, FL, CRC Press, 2019. doi.org/10.1201/b22189
15. Yang Y, Wang C, Zhao H, et al: AI applications to reduce loneliness among older adults: a systematic review of effectiveness and technologies. *Healthcare (Basel)* 2025; 13(5):446. doi:10.3390/healthcare13050446
16. Carel H, Kidd IJ: Epistemic injustice in healthcare: a philosophic analysis. *Med Health Care Philos* 2014; 17(4):529–540. doi:10.1007/s11019-014-9560-2
17. Morley J, Luciano F, Charras K, et al: The ethics of AI in health care: a mapping review. *PLoS One* 2020; 15(9):e0239233. doi:10.1371/journal.pone.0239233
18. Chu CH, Nyrop R, Leslie K, et al: Digital ageism: challenges and opportunities in artificial intelligence for older adults. *Gerontologist* 2022; 62(7):947–955. doi:10.1093/geront/gnab167
19. Budd MA, Mlinac ME: Overview of practical applications of healthcare ethics for older adults. *OBM Geriatrics* 2025; 9(2):314. doi:10.21926/obm.geriatri.2502314

20. Gasteiger N, Loveys K, Law M, et al: Friends from the future: a scoping review of research into robots and computer agents to combat loneliness in older people. *Clin Interv Aging* 2021; 16:941-971. doi:10.2147/CIA.S282709
21. Broadbent E, Loveys K, Ilan G, et al: ElliQ, an AI-driven social robot to alleviate loneliness—progress and lessons learned. *J Aging Res Lifestyle* 2024; 13:22–28. doi:10.14283/jarlife.2024.2
22. Bayen E, Nickels S, Xiong G, et al: Reduction of time on the ground related to real-time video detection of falls in memory care facilities: observational study (SafelyYou Guardian). *J Med Internet Res* 2021; 23(6):e17551. doi:10.2196/17551
23. Eost-Telling C, Yang Y, Norman G, et al: Digital technologies to prevent falls in people living with dementia or mild cognitive impairment: a rapid systematic overview of systematic reviews. *Age Ageing* 2024; 53(1):afad238. doi:10.1093/ageing/afad238
24. Kruse CS, Fohn J, Wilson N, et al: Utilization barriers and medical outcomes commensurate with the use of telemedicine in older adults: systematic review. *JMIR Med Inform* 2020; 8(8):e20359. doi:10.2196/20359
25. Chang F, Bjerkan K, Hvalvik S, et al: Investigating the integration and long-term use of smart speakers by older adults: mixed-methods study. *JMIR mHealth uHealth* 2024; 12:e47472. doi:10.2196/47472

26. Mahmood A, Ackerman M, Jones ML: Voice assistants for health self-management: designing for and with older adults, in Proceedings of the 2025 ACM Conference on Human Factors in Computing Systems (CHI '25) 2025.

doi:10.1145/3706598.3713839

27. Office of the National Coordinator for Health IT (ONC): Health data, technology, and interoperability: certification program updates, algorithm transparency, and information sharing (HTI-1); final rule. Fed Regist. 2024;89(6):1192. Published January 9, 2024. <https://www.federalregister.gov/documents/2024/01/09/2023-28857/health-data-technology-and-interoperability-certification-program-updates-algorithm-transparency-and>. Accessed November 1, 2025.

28. Duggan MJ, Gervase J, Schoenbaum A, et al: Clinician experiences with ambient scribe technology to assist with documentation burden and efficiency. JAMA Netw Open 2025; 8(2):e2460637. doi:10.1001/jamanetworkopen.2024.60637

29. Olson KD, Meeker D, Troup M, et al: Use of ambient AI scribes to reduce administrative burden and professional burnout. JAMA Netw Open 2025; 8(10):e2534976. doi:10.1001/jamanetworkopen.2025.34976

30. Topol E: Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again. New York, NY, Basic Books, 2019. ISBN 9781541644632.

31. Mishkind M, Shore JH, Barrett R, et al: Resource document on best practices in synchronous videoconferencing-based telemental health. *Telemed J E Health* 2024; 30(5):1330-1340. doi:10.1089/tmj.2023.0174
32. Gabbard GO: *Psychodynamic Psychiatry in Clinical Practice*, 5th ed. Washington, DC, APPI, 2014. ISBN 1585624438, 9781585624430
33. Beck JS: *Cognitive Behavior Therapy: Basics and Beyond*, 2nd ed. New York, NY, Guilford, 2011.
34. Charon R: Narrative medicine: a model for empathy, reflection, profession, and trust. *JAMA* 2001; 286(15):1897–1902. doi:10.1001/jama.286.15.1897
35. Chochinov HM: *Dignity Therapy: Final Words for Final Days*. New York, NY, OUP, 2012. ISBN 0195176219, 9780195176216.
36. Forrest CB, Chesley FD Jr, Tregear ML, et al: Development of the learning health system researcher core competencies. *Health Serv Res*. 2018 Aug;53(4):2615-2632. doi: 10.1111/1475-6773.12751
37. U.S. FDA: Marketing submission recommendations for a predetermined change control plan (PCCP) for AI-enabled device software functions. Published Aug 18, 2025. <https://www.fda.gov/media/166704/download>. Accessed November 1, 2025.
38. IEC: IEC 62366-1:2015 Medical devices--usability engineering. Published in 2015 and reviewed and confirmed in 2021.

<https://www.iso.org/obp/ui/en/#iso:std:iec:62366:-1:ed-1:v1:en>. Accessed November 1, 2025.

39. ISO: ISO 14971:2019 Medical devices—Application of risk management to medical devices. Published in 2019 and reviewed and confirmed in 2025.
<https://www.iso.org/standard/72704.html>. Accessed November 1, 2025.

40. The Joint Commission / CHAI. The responsible use of AI in healthcare (RUAIH™). (PDF, ©2025 Joint Commission).
<https://digitalassets.jointcommission.org/api/public/content/dcfcf4f1a0cc45cdb526b3cb034c68c2>. Accessed November 1, 2025.

41. U.S. FDA: National evaluation system for health technology (NEST). Published Oct 29, 2019. (FDA page.) <https://www.fda.gov/about-fda/cdrh-reports/national-evaluation-system-health-technology-nest>. Accessed November 1, 2025.

42. Mitchell M, Wu S, Zaldivar A, et al: Model cards for model reporting. In Proceedings of the Conference on Fairness, Accountability, and Transparency FAT* 2019:220-229. <https://doi.org/10.1145/3287560.3287596>. Accessed November 1, 2025.

43. Gebru T, Morgenstern J, Vecchione B, et al: Datasheets for datasets. Commun ACM 2021; 64(12):86–92. <https://doi.org/10.1145/3458723>. Accessed November 1, 2025.

44. Perez-Toro PA, Dineley J, Iniesta R, et al: Exploring biases related to the use of large language models in a multilingual depression corpus. *Sci Rep* 2025; 15:36197. doi:10.1038/s41598-025-19980-x
45. ONC: 21st century cures act final rule—‘communications’ condition (gag-clause ban). Final rule published May 1, 2020.
46. Rogers CR: The necessary and sufficient conditions of therapeutic personality change. *J Consult Psychol* 1957; 21(2):95–103. doi:10.1037/h0045357
47. Miller WR, Rollnick S: *Motivational Interviewing: Helping People Change*. 3rd ed. New York, NY, Guilford Press, 2012.
48. OpenAI: Helping people when they need it most. Published Aug 26, 2025. (Official post stating safeguards “can be less reliable in long interactions”.) https://openai.com/index/helping-people-when-they-need-it-most/?utm_source=chatgpt.com. Accessed November 1, 2025
49. Kroenke K, Spitzer RL, Williams JBW: The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 2001;16(9):606–613. doi:10.1046/j.1525-1497.2001.016009606.x
50. Spitzer RL, Kroenke K, Williams JBW, et al: The GAD-7. *Arch Intern Med*. 2006;166(10):1092–1097. doi:10.1001/archinte.166.10.1092
51. Jünger S, Payne SA, Brine J, et al: Guidance on Conducting and REporting DElphi Studies (CREDES) in palliative care: recommendations based on a

methodological systematic review. Palliat Med 2017; 31(8):684-706.

doi:10.1177/0269216317690685

Highlights

- What is the primary question addressed by this study?

How can Humane Intelligence, a patient-centered, ethically attuned, clinically grounded relational framework, guide the design, evaluation and use of AI for geropsychiatric care?

- What is the main finding of this study?

This article proposes Humane Intelligence as a four-pillar framework -- Relational Intelligence, Transparency with Care, Reciprocity and Consent, and Ethical Governance in Strategic Regions -- operationalized through a Moral Grid Operational Index with concrete cues and measures. It shows how these pillars translate into point-of-care practices that keep care patient-centered and relational while making AI more transparent, correctable, and safe.

- What is the meaning of the finding?

Applying this framework helps organizations deploy AI that improves patient-important outcomes, trust, and accountability while better protecting older adults.

Box 1: 2025 JAMA Summit on AI Priorities and Humane Intelligence: Structured Insights

Mapping 2025 JAMA AI Summit Priorities to the Four Pillars of Humane Intelligence

The 2025 JAMA AI Summit identified four system-level priorities for advancing artificial intelligence in healthcare. These priorities are strategically aligned with the Four Pillars of Humane Intelligence to guide their practical adoption and implementation across the field.

- **Multistakeholder Total Product Lifecycle (TPLC) Management:** This priority emphasizes the need for inclusive management across the full lifecycle of healthcare products. It aligns with Humane Intelligence Pillar 1 (Relational Intelligence) and Pillar 3 (Reciprocity and Consent), promoting collaborative relationships and respect for individual autonomy throughout the development and deployment process.
- **Tools for Evaluation and Monitoring:** Robust evaluation and monitoring tools -- such as APPRAISE-AI, CONSORT-AI, DECIDE-AI, and TRIPOD+AI -- are essential not only for safety, but also for effectiveness. This priority maps to Humane Intelligence Pillar 2 (Transparency with Care), ensuring that AI systems are evaluated transparently and responsibly.
- **National Learning Data Infrastructure:** Building representative and federated learning data infrastructures, capable of rapid causal evaluation, is vital for national progress. This priority is linked to Humane Intelligence Pillar 4 (Ethical Governance in Strategic Regions), underscoring the importance of ethical stewardship in data management and usage.

- Aligned Incentives: Creating policy and market incentives to enhance feasibility is also mapped to Humane Intelligence Pillar 4 (Ethical Governance in Strategic Regions), encouraging responsible and sustainable AI adoption through regulatory and economic alignment.

Box 2: The CASR Procurement Checklist

The Criteria-Accountability-Support-Recourse (CASR) Procurement Checklist provides a structured approach for evaluating AI procurement in healthcare, ensuring transparency, accountability, support, and recourse.

- **Criteria:** Clearly define success metrics—such as clinical impact, burden reduction, and equity—and establish thresholds for performance (e.g., reducing 30- and 90-day emergency department revisits by at least 10% without increasing subgroup disparities).
- **Accountability:** Assign explicit ownership for model, data, consent, and escalation processes (e.g., a designated program owner who approves quarterly change-logs and bias checks).
- **Support:** Ensure that resources are available for training, supervision, and accessibility (e.g., training 20 clinicians, providing large-font user interfaces, and offering captioned instructional videos for families).
- **Recourse:** Implement mechanisms for stopping, undoing, explaining, and redressing errors, supported by Service Level Agreements (SLAs) and tracking (e.g., reversing a mistaken discharge readiness flag within 24 hours, providing an explanation within 3 days, and resolving the issue within 14 days).

This procurement checklist is consistent with concepts from ONC HTI-1 DSI transparency, FDA PCCP change control, and EU AI Act documentation requirements for high-risk health applications. These references serve as guidelines and should be adapted to comply with local regulatory frameworks.

Box 3: Routing Rules and Escalation Patterns for General-Purpose Chatbots:

“Not for Crisis”

General-purpose chatbots are not designed, tested, or approved for use as crisis intervention tools and should not be considered primary resources in situations involving imminent or escalating risk to individuals.[2]

- Clearly disclose any limitations regarding crisis intervention up front.
- Rapidly route acute situations to human assistance (such as the 988 hotline in the United States or other international crisis services).
- Utilize licensed clinical applications with human escalation pathways for assessment and treatment workflows.
- Strictly avoid content that describes harmful methods.
- Recognize that safeguards may degrade in emotionally charged or prolonged exchanges, increasing real-world risks.[48]
- Adhere to incident reporting and postmarket guidance when harm occurs.[2]

Routine AI safety measures focus on reliability, fairness and bias moderation, robustness, privacy and security, and explainability, usually within a risk-management framework. Additional safeguards include independent model cards (evaluation records and logs), subgroup monitoring with rollback triggers, public incident reporting, third-party audits, and vendor conformance terms.

Policy anchor: Fully automated clinical actions are not permitted. AI systems may inform, summarize, and forecast, but should not order, prescribe, discharge, detain, or deny care without a responsible human clinician verifying and signing off on these decisions.

Geriatric “red-team” testing and formal change control for updates are essential for older adult safety.

Safety also depends on defining the AI scope-of-action—clarifying roles, responsibilities, and limits. These include, progressively increasing in complexity: assistive actions, advisory actions, guided collaborative production (with clinician oversight), and autonomous clinical action (which is not recommended under Humane Intelligence principles). Each advancement requires stronger evidence, more stringent safeguards, and explicit ongoing consent.

Box 4: Governance as Caring Posture: Program Patterns for Older Adult Safety and Dignity

This section provides a checklist for service chiefs, quality improvement leads, and other program leaders evaluating new AI tools intended for older adults.

- Develop specialized programs rather than generic chatbots (e.g., ML-Lite for Older Adults, Person-Centered Therapy (PCT) -- Companion Mode, Medication Review Explainer, Fraud and Scam Sentinel, Falls-Risk Check-In).
- Include robust guardrails: hard constraints, no-advice zones, senior-friendly language, ethical refusals, and immediate human hand-off criteria.
- Promote uncertainty disclosure through user interfaces that display confidence levels and underlying explanations (such as data class, recency, and consensus versus controversy).
- Advance data dignity with explicit, revocable consent, local preference memory with editable logs, and a strict opt-in policy for secondary data use.
- Assess outcomes relevant to older adults, including engagement, comprehension, decisional conflict, safety events, and equity measures; conduct subgroup analyses for this population.
- Design for accessibility, considering social determinants of health (such as rural or low-bandwidth environments, offline fallback options, and multilingual access for older adults and caregivers).

These program patterns help care teams differentiate between generic chatbots and AI solutions tailored to the unique needs of older adults, thereby enhancing safety, dignity, and effectiveness in care delivery.

Table 1. Moral Grid Operational Index and Ethical Governance in Clinical AI

Moral Grid Operational Index: Point-of-Care Behaviors and Auditable Evidence

The four pillars of Humane Intelligence are the organizing structure; the Moral Grid Operational Index operationalizes them through cross-cutting dimensions (e.g., presence, equity, safety/supervision, data stewardship) and provides a clear linkage to specific, observable behaviors at the point-of-care, alongside concrete, auditable evidence. This structure ensures that ethical practices are embedded within clinical workflows and can be reviewed systematically. Ethical actions are transparent and reviewable at the point-of-care.

- Relational Intelligence: Presence-related behaviors: Clinicians perform attuned introductions, transparently disclosing roles and intentions. Evidence for this includes signed or recorded consent and session logs.
- Transparency with Care: At every visit, clinicians recap the AI's role and limitations in plain language. This is documented through versioned evaluation records, logs with references, and notes on changes.
- Reciprocity and Consent: Dignity, autonomy, consent integrity: Patient autonomy is respected, with all stop or undo requests honored and mediated consent procedures enacted when necessary. Evidence includes tracking of refusal events and fulfillment of recourse.
- Ethical Governance in Strategic Regions:
 - Access, inclusion, cost transparency, and equity delta* monitoring: Outcomes are stratified by age, language, and disability to monitor

disparities. Dashboards check for equity delta changes and log action items for follow-up.

- Safety and supervision: When uncertainty is flagged, escalation protocols are triggered. Audit logs capture escalation reasons and outcomes.
- Data Stewardship: Data use is limited to the minimum necessary, with patient portal access provided. Evidence includes a data map and a defined retention schedule.

*Equity delta refers to before-and-after changes in disparity.

Table 2. Moral Grid Operational Index: Risk-Tiered Safeguards and Operational Cues

Safeguards are tiered by risk level (T1-T3), with operational cues for uncertainty, consent, equity, and audit. The table below details how practices and escalation procedures scale with risk, ensuring appropriate ethical oversight.

	Routine, Low Risk (Tier 1; T1)	Sensitive, Moderate Risk (Tiers 1-2; T1–T2)	High-Stakes (Tiers 2-3; T2–T3)	When to Escalate to Ethics Review Committee
Relational Intelligence	Recognize the individual; assist quietly; document learning and value trade-off discussions.	Surface values, invite preferences, explain options, and document value trade-off discussions.	Explain trade-offs explicitly, ensure support, plan follow-up, and document value trade-off discussions.	Defer irreversible decisions, involve advocacy, slow the pace, and document value trade-off discussions.
Transparency with Care (State uncertainty explicitly)	Plain-language notes; avoid unnecessary exposure.	Check understanding; share uncertainties and limitations.	Provide rationale and safeguards; confirm comprehension.	Disclose conflicts; pause for consent or committee consultation.

Reciprocity and Consent	Affirm consent for routine support and obtain consent receipt.	Renew consent and ask about current priorities; obtain consent receipt.	Re-obtain informed consent and offer alternatives; obtain consent receipt.	Escalate to ethics review committee when consent is unclear or contested; obtain consent receipt.
Ethical Governance in Strategic Regions	Light-touch logging and periodic review; log equity delta (=before and after changes in disparity) checks.	Risk-aware logging and supervisor spot-checks; log equity delta checks.	Full audit trail, attending oversight, clear handoffs; log equity delta checks.	Obtain multidisciplinary consultation, document decisions, and maintain learning loops; log equity delta checks.

Graduated autonomy and risk (T1–T3) are matched to safeguards such as uncertainty alerts, consent renewal, equity delta monitoring, and audit depth. Ethics review committee consultation is triggered when consent is unclear or disputed. Alert flags for equity deltas can signal patterns of uneven benefit or burden. These frameworks empower clinicians, ethicists, and caregivers to evaluate and design emerging AI systems in healthcare.

Table 3. SUNSHINE and Humane Intelligence: Bridging Systems to Point-of-Care Practices

SUNSHINE system components are linked to Humane Intelligence concepts at the point-of-care, with metrics and evidence examples demonstrating the integration of system-level design and frontline practice.

SUNSHINE component (systems view)	Humane Intelligence concept (point-of-care practice)	Metric/evidence examples
Integrated data and AI/ML analytics across settings	Transparency with Care evaluation records, model cards (records/logs), and “why-this-alert” explanations at point-of- care	Percentage of encounters with uncertainty badge/indicator; time-to- understanding survey
Care coordination and transitions	Relational Intelligence prompts to surface values and meaning, and continuity cues for teams	Documented values/preferences attached to orders; continuity index across transitions
Family/caregiver engagement	Reciprocity and Consent as living consent, including opt-downs, preferences and consent updates	Consent update cadence; percentage of caregiver- acknowledged preferences honored

Public health and community partnerships	Ethical Governance in Strategic Regions with equity targets and vendor accountability	Moral Grid Operational Index by region (bias, access, outcomes); public audit frequency
Evaluation of resilience outcomes	Moral Grid Operational Index linking outcomes to safety, bias, explainability, and consent guardrails	Pre-specified failure-mode logs; disparity reduction over baseline; post-event disclosure rate

These bridges between systems and point-of-care practices are supported by concrete metrics and evidence, such as explanation rates, consent updates, and public audit frequencies.

Table 4. Outcomes and Measurement Hooks for Humane Intelligence Ethical Governance in Strategic Regions

Outcomes and measurement hooks are defined for ethical governance, using familiar clinical measures and clear accountability nodes. The table below outlines the approach for each outcome.

Outcome	Example measure(s) familiar to clinicians	Unit of analysis	Collection cadence	Trigger/threshold for review*	Accountability node (Moral Grid)
Function	ADLs/IADLs ; WHODAS 2.0; PROMIS-Physical Function	Patient/service	Per encounter; monthly roll-up	≥10% decline in function over 30 days among AI-exposed vs baseline or matched non-exposed	Point-of-care Lead → Program Owner; Quality and Safety Board if trend persists
Distress	GAD-7; PHQ-9 or GDS; CSDD (for those with major	Patient	Per encounter (screen); weekly review	PHQ-9 increase ≥5 or suicidality item >0; CSDD increase ≥4;	Clinician-in-Charge (rapid response) → Program Owner

	neurocognitive disorder); NPRS (for pain)			acute risk → immediate human review	
Caregiver burden	Zarit Burden Interview (ZBI-12); Caregiver Strain Index (CSI)	Dyad (patient–caregiver)	Monthly	ZBI-12 increase ≥ 6 or score ≥ 20 for 2 cycles	Care Team Lead → Patient/Caregiver Council (feedback loop)
Avoidable utilization	ED visits; 30-day readmissions; unplanned admissions; polypharmacy flags	Service/site/system	Monthly; quarterly audit	$\geq 15\%$ increase vs 3-month baseline among AI-exposed; readmits > benchmark	Program Owner → Quality and Safety Board; Payer/Utilization Partner
Equity deltas (=before and after	Stratified performance (age bands,	Site / system	Quarterly	$\geq 20\%$ relative disparity in any outcome above (after	Equity and Ethics Panel → Executive

changes in disparity)	race/ethnicity, language, cognitive status, disability, insurance)			case-mix adjustment)	Sponsor; corrective action logged
Document ed harms / overrides	Adverse events attributed to AI advice; near-miss reports; clinician override rate; “do- not-follow” flags	Program	Weekly review; monthly summary	Any serious adverse event; near- miss cluster; override rate >30% in a month	Program Owner → Independent Safety Reviewer; vendor notified

*Thresholds are examples; each site sets its own values and publishes case-mix rationale.

Rapid triggers (such as suicidality or serious adverse events) are routed immediately to the clinician-in-charge.

Accountability nodes refer to accountable roles of entities that can act when a trigger or threshold for review occurs, and includes roles such as the Point-of-care Lead, Program

Owner, Quality and Safety Board, Patient/Caregiver Council, Equity and Ethics Panel,
Executive Sponsor, and Independent Safety Reviewer.

Table 5. Clinical Safety Language: Actionable Failure Tags for Detection, Mitigation, and Documentation

A standardized safety vocabulary supports structured logging of AI-related failure modes, detection, and mitigation at a local incident level. The following tags and practices facilitate learning loops and audits:

Failure type (Notation)	Typical Trigger Example*	Detection Methods	Mitigation and Documentation Goals
Fabricated content (FAB) (Unsupported assertion)	Clinical or quantitative claim with no retrievable source.	Source-checking; retrieval traces	Enforce source/confidence thresholds; tag “FAB”; include source links in notes.
Misattribution (MISATTR) (Wrong source/identity)	Citation text does not match the linked article.	Citation entity linking; DOI/PMID/URL cross-check	Stricter attribution prompts; tag “MISATTR”; correct source in notes.
Omission (OMIT) (Critical context excluded)	High-relevance factor (e.g., allergy, drug level) documented	Content checklists by topic; coverage prompts to confirm/negate key factors	Trigger context re- ask; tag “OMIT”; add missing context and re- issue note.

	elsewhere is missing.		
Decontextualization (DECON) (Answer not fit to setting)	Setting/profile mismatch (e.g., inpatient is given ambulatory guidance to go to the ED).	Setting/profile mismatch tests; routing rules by setting	Profile-aware routing; tag “DECON”; add rationale or correct context.
Overreach (OVER) (Speculation presented as fact)	Speculation presented as fact without adequate evidence or uncertainty language.	Hedge-vs-claim detector; scope/policy hard-stops for high-stakes claims	Force uncertainty statements; redirect out-of-scope items; tag “OVER”; include uncertainty line.
Calibration error (CAL) (Confidence-reliability mismatch)	Over/under-estimation of effect with sparse/low quality evidence or historically low accuracy on similar items.	Reliability diagrams; Brier score** decomposition	Post-hoc calibration; tag “CAL”; display level of confidence adjusted to known reliability in notes.
Agreeableness/deference	Model echoes	Contradiction	Reward accuracy

bias (AGREE) (Mirrors stated views)	incorrect claim without evidence or counter- example.	probes; stance detection against vetted sources; counter-example retrieval	over agreement; tag “AGREE”; require evidence or guideline citation.
--	--	--	---

*Sites may refine trigger values and thresholds to their local policy.

**The Brier score is a probability-error metric; its reliability component quantifies mismatch between stated confidence and observed outcomes, with lower scores indicating higher accuracy.