Beyond the Hospital: A Residential Care Environment Simulator for Evaluating AI in Long-Term Care Facilities

Riccardo Crosa

Montecristo International OÜ, Tallinn, Estonia

Università del Piemonte Orientale, Facoltà di Scienze Politiche (Laurea in Servizi alle Imprese e alle Organizzazioni, 2005). President, Nord Servizi Società Cooperativa Sociale; Sole Administrator, Alyssum Srl.

Correspondence: riccardo.crosa@montecristo.it

Clinical evaluation frameworks for artificial intelligence have made a decisive leap forward. The Clinical Environment Simulator (CES) proposed by Luo and colleagues in *Nature Medicine* replaces static benchmarks with dynamic hospital simulations where every AI decision reshapes the evolving clinical landscape.[1] Their architecture — a hospital engine tracking bed availability, staffing, and equipment alongside a patient engine modelling disease progression — captures cascading consequences that isolated test cases cannot. This is a genuine advance. But it carries an implicit assumption that deserves scrutiny: that the hospital is the default environment for clinical AI, and that acute care temporality is the relevant timescale for evaluation.

For the estimated 30–40 million people living in residential long-term care facilities worldwide — including over 300,000 older adults in Italy's Residenze Sanitarie Assistenziali (RSA) system alone[2,3] — this assumption excludes the very setting where AI deployment may prove most consequential and most ethically fraught. As AI tools proliferate in residential care — fall prediction algorithms, medication management systems, ambient monitoring, conversational agents for residents with dementia[4,5] — the field lacks any systematic framework for evaluating whether these systems perform safely, effectively, and appropriately within the distinctive ecology of long-term care.

This paper extends the theoretical framework of *organizational recapitulation*[11] — the hypothesis that multi-agent AI systems structurally reproduce classical organizational dynamics by structural necessity — to the specific domain of residential elder care. Where that framework establishes the general theory, the Residential Care Environment Simulator (RCES) proposed here provides its first domain-specific application: an evaluation paradigm for what may become one of the most consequential areas of AI deployment in healthcare.

Why hospitals are not the right model

Long-term care facilities differ from hospitals in almost every dimension that matters for AI evaluation. The temporal horizon extends from days to years. The clinical picture is dominated by multi-morbidity, polypharmacy, and progressive functional decline rather than discrete diagnostic episodes. The workforce is predominantly composed of care assistants and nurses rather than

physicians — in Italy, a typical RSA with 60 beds may have a physician present only a few hours per week, while care assistants provide 24-hour coverage at ratios that make continuous clinical triage a structural necessity rather than an emergency protocol.[6] Documentation remains largely paper-based in many facilities, and the institutional logic must balance medical imperatives with social, emotional, regulatory, and relational demands that have no parallel in acute care.

I write from direct experience. As president of Nord Servizi Società Cooperativa Sociale and sole administrator of Alyssum Srl — a holding company owning three RSA properties serving approximately 150 residents — and simultaneously as founder of a near-zero human company where 95% of operations are managed by AI agents, I occupy an unusual vantage point at the intersection of residential elder care and applied artificial intelligence. The facilities operate under a private management model developed over 26 years by Giovanni Battista Caprara, which demonstrated that cooperative-owned residential care could achieve both clinical quality and financial sustainability outside the rigid institutional frameworks inherited from Italy's former IPAB system[12] — a model that now provides the operational substrate on which AI evaluation must be built. Operational management is delegated to Cooperativa Sociale Quadrifoglio (Pinerolo), one of Italy's largest third-sector care providers with approximately 4,000 workers and €140 million in annual revenue.[7]

The operational reality I observe daily — chronic understaffing, paper-based records, the slow accumulation of subtle clinical signals across months of residency, the irreducible importance of staff-resident relationships built over years — is invisible to evaluation frameworks designed for the compressed temporality of hospital care.

The Residential Care Environment Simulator

I propose the Residential Care Environment Simulator (RCES), a conceptual framework that adapts and fundamentally extends the CES paradigm for long-term care. The RCES comprises three interacting engines, each capturing a dimension of residential care that is absent or marginal in hospital-oriented frameworks.

The **facility engine** models the operational infrastructure of a residential care facility: shift-based staffing patterns with chronic understaffing as the norm; the ratio of care assistants to nurses to visiting physicians; supply logistics for medications, continence products, and wound-care materials; regulatory compliance workflows including periodic assessments mandated by national frameworks; and the physical environment — shared rooms, communal dining areas, gardens — that shapes both clinical and social outcomes. Crucially, the facility engine models institutional memory: in residential care, staff members accumulate deep knowledge of individual residents over months and years, knowledge that AI systems must learn to complement rather than displace.

The **resident engine** simulates the health trajectories of long-term care populations. Where hospital patient engines model acute disease progression, the resident engine captures what geriatricians call the dwindling trajectory: slow, non-linear decline punctuated by acute episodes that may or may not resolve to a new, lower baseline. It models multi-morbidity as the default state — residents typically carry five to eight concurrent chronic conditions whose interactions generate

emergent complexity. It simulates polypharmacy cascades, tracks functional status across activities of daily living, cognitive capacity along validated decline trajectories, nutritional status, skin integrity, continence, mood, and social engagement. And it models end-of-life transitions as a dynamic process rather than a binary switch, reflecting the clinical reality that terminal trajectories are often ambiguous and contested.

The **relational engine** — the component with no analogue in hospital-oriented frameworks — captures the social and emotional architecture of residential care. Residents are not patients passing through; they are people living in the facility, often for years, building relationships with staff and co-residents while maintaining or losing connections with family. The relational engine models family dynamics, staff-resident attachment, social networks among residents, and the regulatory environment of advance directives, capacity assessments, and complaint mechanisms. An AI system that correctly identifies a medication interaction but communicates its recommendation in a way that disrupts the trust between a care assistant and a resident has failed in a way that no clinical benchmark would capture.

Three evaluation capabilities absent from current benchmarks

The RCES enables three critical evaluations that no existing framework measures. First, **longitudinal coherence**: can an AI system recognise that a resident's increasing agitation is not a new behavioural disturbance but the cumulative effect of a medication change six weeks ago? Can it detect the subtle inflection point signalling a transition from chronic management to end-of-life care? The RCES presents AI systems with resident trajectories spanning months of simulated time, requiring integration of dispersed signals across fragmented documentation.

Second, **resource allocation under chronic scarcity**: where hospital simulators model resource constraints as acute crises, the RCES models scarcity as a permanent structural condition. Which residents receive the most attention from limited nursing staff? How should a visiting physician's few weekly hours be allocated across dozens of competing needs? An AI system that consistently recommends specialist referrals in rural areas with no geriatrician within 100 kilometres, or one-to-one supervision that the staffing ratio cannot support, is failing a test that no current benchmark measures.

Third, **ethical navigation**: long-term care ethics involves sustained navigation of resident autonomy in the context of cognitive decline — the right to refuse medication, to eat foods that increase aspiration risk, to decline physiotherapy — balanced against duty-of-care obligations, mediated between residents' wishes and families' preferences, and shaped by cultural frameworks around ageing, dependency, and death. The RCES evaluates whether AI systems can recognise ethical complexity, flag decisions requiring human deliberation, and avoid both paternalistic override and negligent deference.

A call for slow AI

The RCES framework proposes that AI performance in long-term care be scored across four interconnected domains: clinical outcomes (medication safety, avoidable hospitalisations, fall prevention), quality-of-life indicators (functional independence, social participation, dignity), regulatory compliance, and relational integrity. AI systems that optimise clinical safety at the expense of quality of life — recommending physical restraints to prevent falls, or flagging every medication refusal for capacity assessment — would score poorly.

This framework also exposes the severe data gap limiting AI development for long-term care. Hospital electronic health records have generated enormous training datasets; residential care documentation remains fragmented, inconsistently structured, and often paper-based. Building the RCES will require systematic collection of longitudinal functional trajectories, relational histories, and quality-of-life measurements that most facilities do not currently capture in structured form.

Ultimately, the deployment of AI in residential care represents a test case for what I would call *slow AI* — systems designed not for the rapid, high-stakes decisions of emergency medicine but for the sustained, relationship-embedded, ethically complex work of caring for people over the long arc of ageing and dying. If we are serious about evaluating such systems, we need tools as patient and attentive to human complexity as the care they are meant to support. The RCES is a first step toward building them.

Declaration of interests

RC is CEO of Montecristo International OÜ (Estonia), president of Nord Servizi Società Cooperativa Sociale, and sole administrator of Alyssum Srl, a holding company owning three RSA properties in Italy. The operational management of these facilities is delegated to Cooperativa Sociale Quadrifoglio (Pinerolo, Italy). The reference shareholder of Alyssum Srl is Giovanni Battista Caprara. RC reports no other conflicts of interest.

References

1 Luo L, Kim SE, Zhang X, et al. A clinical environment simulator for dynamic AI evaluation. *Nat Med* 2026; **32**: 820–27.

2 ISTAT. Presidi residenziali socio-assistenziali e socio-sanitari: anno 2022. Rome: Istituto Nazionale di Statistica, 2023.

3 United Nations, Department of Economic and Social Affairs. World Population Ageing 2023: Highlights. New York: UN, 2023.

4 Bauer A, Strudwick G, Engel L, et al. Artificial intelligence for older people receiving long-term care: a systematic review. *Lancet Healthy Longev* 2022; **3**: e286–97.

5 Khalil RA, Ahmad K, Ali H. Redefining elderly care with agentic AI: challenges and opportunities. *arXiv* 2025; 2507.14912.

6 OECD. Health at a Glance 2023: OECD Indicators. Paris: OECD Publishing, 2023.

7 Cooperativa Sociale Quadrifoglio. Bilancio sociale al 31/12/2023. Pinerolo, 2024.

8 Artificially intelligent nursing homes: a scoping review of palliative care interventions. *Front Digit Health* 2025; **7**: 1484304.

9 Bean DM, Stringer C, Beeknoo N, et al. LLM-assisted self-triage in the general population. *Nat Med* 2026; published online.

10 Gerke S, Babic B, Evgeniou T, Cohen IG. Autonomous AI for prescription renewal. *N Engl J Med* 2026; published online.

11 Crosa R. Organizational recapitulation: toward a sociological theory of multi-agent AI systems. Preprint, Montecristo International OÜ, 2026. Available at mci.ee/recapitulation. Submitted to *Sociological Theory*.

12 Caprara GB. Private communication. The Caprara model of cooperative-owned RSA management, developed 1998–2024, demonstrated sustainable alternatives to IPAB-derived institutional frameworks across multiple facilities in Northern Italy.