Universal Library of Innovative Research and Studies Research Article

ISSN: 3065-0003 | Volume 3, Issue 4

Open Access | PP: 20-26

DOI: https://doi.org/10.70315/uloap.ulirs.2026.0304004

The Care Plan Compiler as an Agentic AI Framework for Translating Clinical Recommendations into Feasible Patient-Centered Care

Sakshit Teotia

Principal AI Tech Architect, New York, USA.

Abstract

A clinical recommendation becomes care after someone confirms coverage, finds a facility that provides the service, arranges language support and transportation, books an appointment, and verifies that the visit took place. When no organization owns this work, it falls to the patient, and a missed service is recorded as nonadherence, even when the failure occurred within the health system. The review examines five adjacent fields, each addressing part of this problem, and develops the Care Plan Compiler, an agentic AI framework that treats the clinician-approved recommendation as a protected source specification and compiles it into a graph of tasks with named owners. A deterministic readiness checker releases a task once all prerequisites hold and classifies any unknown coverage, capacity, or accessibility as unresolved. Feasible plans are compared across a multidimensional burden vector, and a change in context triggers recompilation of the dependent tasks. A synthetic case of diabetic retinal screening illustrates the compilation, and a four-stage evaluation program sets a screening completed within the approved window and verified within 90 days of referral approval as its primary endpoint. The framework moves responsibility for missed care to the system layer where the failure occurs.

Keywords: *Agentic AI, Care Planning, Treatment Burden, Patient Navigation, Clinical Workflow, Human Oversight, Health Equity.*

INTRODUCTION

A referral for diabetic retinal screening fits on one line of a clinical note, yet carrying it out takes a chain of operational steps that includes coverage and authorization, a facility that performs the test, language support, an appointment, transport, records, and proof that the examination took place. In the United States, patients and their families perform much of this administrative work through forms, calls, and coordination among payers and providers (Kyle & Frakt, 2021).

Losses accumulate at each transition of the chain. In a safety-net system in the Bronx, 35.1% of patients referred to a retina specialist after a positive teleretinal screen completed the visit within 1 year (Dahlan et al., 2023). Among patients with limited English proficiency, 53% of stool DNA tests produced no usable result, compared with 29% among English-proficient patients (Hill et al., 2022). A record that labels such outcomes as nonadherence places the failure on the patient, although the cause may lie in an absent service, a missing authorization, an inaccessible facility, or a task that no organization accepted. First-hand work on the modernization of a state Medicaid eligibility and renewal process showed the same displacement on the administrative side, where direct analysis of unresolved cases traced lost coverage to systemic errors and to tasks without a clear owner.

Agentic AI systems can hold a goal, call software tools, observe results, and revise a plan, which opens this chain to automation and to new risks, since an agent may act on stale availability data, disclose more than a task requires, or alter a clinical order while simplifying logistics. A scoping review of 43 studies on healthcare AI agents found that most evaluations occurred in simulated or laboratory settings, with limited reporting of safety endpoints (Njei et al., 2026). Across evaluations of large language models in health care, 5% of studies used real patient care data (Bedi et al., 2025). Direct design work on agentic contact-center systems for a health insurer pointed to a workable boundary, in which virtual agents resolve routine eligibility and benefits questions and route every case beyond that scope to human staff.

The review maps what adjacent fields contribute to the path from recommendation to completed care and develops the Care Plan Compiler, an agentic AI framework built on one proposition: a care plan is ready for execution when each task is linked to approved clinical intent, assigned to a responsible actor, and checked against the current circumstances of the patient and the operating conditions of the health system.

Citation: Sakshit Teotia, "The Care Plan Compiler as an Agentic AI Framework for Translating Clinical Recommendations into Feasible Patient-Centered Care", Universal Library of Innovative Research and Studies, 2026; 3(4): 20-26. DOI: https://doi.org/10.70315/uloap.ulirs.2026.0304004.

METHODS

The study combined a targeted literature review with the conceptual design of an architecture. The review covered peer-reviewed publications from 2021 to 2026 indexed in Crossref and PubMed, documentation of interoperability standards, and public descriptions of care-planning prototypes released through October 2026. Search terms paired agentic AI and large language models with care planning, referral completion, patient navigation, administrative burden, and social context, and exact-phrase searches tested whether the construct had appeared under a similar name.

Each approach was coded against eight capabilities: representation of clinical intent, protection of that intent from modification by nonclinical actors, use of patient context, assignment of an actor to each task, a readiness gate before task release, treatment of unknown facts as unresolved, comparison of options by burden, and replanning after a change in context. A capability is counted as present when the source described it as a design property and as partial when the source described a narrower form. Design principles, a formal specification, and a synthetic case of diabetic retinal screening followed from the gaps, and the evaluation program adopted the staged logic of DECIDE-AI (Vasey et al., 2022). The search was targeted and did not cover every database, the worked example uses a synthetic patient, and the framework has not yet been implemented or tested on patient data.

RESULTS

Contributions of Adjacent Fields

Computable clinical guidance forms the first field. Guideline execution engines, the HL7 FHIR PlanDefinition, CarePlan, and Task resources (HL7 International, 2023), and the WHO SMART Guidelines (World Health Organization, n.d.) encode what clinicians intend and when an action applies. Guideline engines check clinical preconditions and react to new clinical data, and FHIR resources can name a task owner, which Table 1 records as partial capabilities, yet none of these tools ask whether the resulting work is feasible for a particular person at a particular time.

Research on treatment and administrative burden forms the second field. It frames forms, calls, travel, and coordination as a cost that the health system transfers to patients (Kyle & Frakt, 2021) and shows that language barriers deepen that cost (Hill et al., 2022), while measuring burden after care has been planned.

Social-context data forms the third field. Models can surface transport needs, work schedules, or caregiving duties from clinical notes, but people with limited access to care have less reliable electronic health records, which degrade the AI models trained on them (Zink et al., 2026). An extracted fact remains a hypothesis until someone confirms it.

Human patient navigation forms the fourth field and the closest functional analog of the framework. A meta-analysis of randomized trials found that navigation raised breast cancer screening with a risk ratio of 1.50 (Nelson et al., 2025), and in a randomized trial of community eye screening, 51.8% of participants with a navigator attended the follow-up examination, compared with 38.1% under usual care (Hark et al., 2025). The logic of navigation rests with individual staff members' judgment, making it hard to audit and scale.

Agentic AI forms the fifth field. Care-planning agents generate and debate candidate plans under human review, administrative agents operate payer portals and record systems, and their evaluations center on efficiency and diagnostic accuracy (Njei et al., 2026). Two prototypes released in 2026 come closest to the present proposal: one testing whether combined care plans fit a patient's weekly capacity (Abdul Salam, 2026) and the other enforcing deterministic policies on agents that coordinate dementia care (Medda & Gong, 2026). Neither assigns every task to an accountable actor, treats unknown prerequisites as blocking, or closes the loop with verified completion.

Comparison of Approaches

Table 1 summarizes the coding of each approach against the eight capabilities.

Table 1. Capabilities of adjacent approaches relative to the Care Plan Compiler

Approach	Clinical intent	Intent guard	Patient context	Actor per task	Readiness gate	Unknown blocks	Burden vector	Replans
Guideline execution engines	Yes	Partly	Partly	No	Partly	No	No	Partly
HL7 FHIR resources	Yes	No	Partly	Partly	No	No	No	No
WHO SMART Guidelines	Yes	Partly	No	No	No	No	No	No
Burden research	No	No	Yes	No	No	No	Partly	No
Social-context extraction	No	No	Partly	No	No	No	No	No
Human patient navigation	Partly	Yes	Yes	Partly	Partly	Partly	Partly	Partly
Care-planning agents	Partly	Partly	Partly	Partly	No	No	No	No
Administrative agents	No	No	No	Partly	No	No	No	No
Capacity-aware prototype, 2026	Yes	Yes	Yes	No	Partly	No	Partly	Partly
Governed dementia-care agents, 2026	Partly	Yes	Partly	Partly	Partly	No	No	Partly
Care Plan Compiler	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Standards represent intent, burden research and navigation describe the operational work, and agents execute steps, yet within the reviewed literature, no approach gates the release of actor-assigned tasks on a deterministic check that treats unknown facts as unresolved. The Care Plan Compiler targets this combination.

Architecture of the Care Plan Compiler

The framework separates three forms of authority. Clinicians approve clinical intent, patients set preferences, consent, and acceptable trade-offs, and the system organizes work and performs bounded actions for which it holds explicit permission. Nine design principles follow from this separation. Approved goals, actions, alternatives, timing, and monitoring requirements are versioned and closed to nonclinical agents, while the system may add enabling work such as coverage inquiries or transport coordination when each task traces back to the approved objective. An inference becomes a fact after confirmation by the patient, an authoritative system, or a professional, and unknown coverage, capacity, or accessibility never converts into a positive claim. Every task has an actor, so the patient does not become the default owner of organizational work. Burden has several dimensions that the system keeps apart, and the absence of a feasible plan counts as a valid result. A change in a supporting fact invalidates the tasks that depend on it. The ninth principle ties the level of human control to the consequence of an action, and the confidence of a model plays no part in that decision. Fig. 1 maps these principles onto the components of the framework.

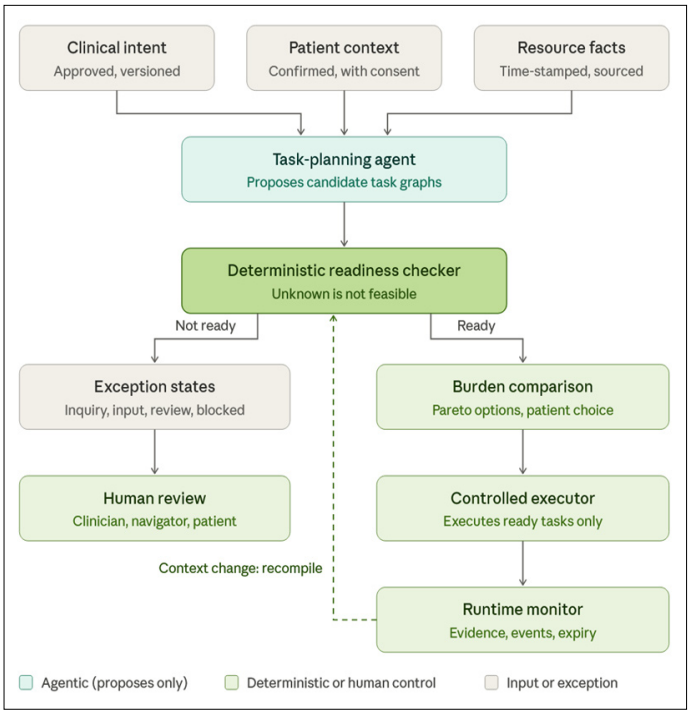


Fig. 1. Architecture of the Care Plan Compiler

Three data inputs enter the compiler under an applicable policy set: approved clinical intent, confirmed patient context, and time-stamped resource facts. A task-planning agent proposes candidate task graphs and has no authority to mark its own tasks ready. A deterministic readiness checker sends a task that fails a predicate to an exception state, which opens an inquiry task for a missing fact or calls for human review of a decision. Candidate plans built from tasks that pass the check go to burden comparison and patient choice. A controlled executor, the single component holding external credentials, performs ready tasks, and a runtime monitor collects completion evidence and returns dependent tasks for recompilation after a change in context.

The protected source specification Σ holds seven elements: goals G, required actions A, approved alternatives O, hard clinical constraints H, urgency and timing U, monitoring requirements M, and version V. The context X_t holds confirmed facts, preferences, known constraints, consent scope, and resource observations, each with a source and an observation time. The compilation function C maps Σ, X_t, and the policy set Π_t into a task graph T_t and a set of structured exceptions E_t, and an empty graph with explicit exceptions counts as a correct output. A task becomes ready when nine predicates hold at once: traceability to approved intent, an authorized actor, consent, an available resource, permitted timing, language and accessibility fit, financial compatibility, policy compliance, and completed dependencies. Under three-valued logic, an unknown value fails a requirement for true, and an unknown prerequisite justifies an inquiry task alone. In addition to being ready, a task can have nine additional statuses. The checker assigns inquiry required, patient input required, professional review required, blocked, and no feasible option, the patient can decline a task, and the runtime monitor records the states expired, completed, and failed.

Burden enters the framework as a nine-dimensional vector: patient time, out-of-pocket cost, travel, contacts, forms, digital steps, cognitive effort, physical effort, and caregiver work. One plan dominates another when it is no worse on every dimension and better on at least one, and the compiler shows the patient the non-dominated options, since no universal formula ranks time against cost or privacy. Each task declares the facts on which it depends, so an expired authorization invalidates scheduling tasks, and a changed work schedule invalidates only the time-dependent options. Every exception carries a code that names the failing layer, from coverage and service capacity to transport, ownership, or patient decision, which turns a missed service into a record of a specific system state.

Consequence tiers decide which ready tasks may run on deterministic checks alone. Each action type receives one of four tiers at deployment, and the planner cannot change that assignment. Tier 0 covers internal read-only actions such as retrieving a coverage record, querying a facility directory, or computing a burden vector. Tier 1 covers internal reversible writes such as creating a task record or logging an inquiry, and

these actions also run on deterministic checks, with an audit entry for each. Tier 2 covers external contact or disclosure, for example a message to the patient, a request to a facility, or the transfer of minimum necessary data to a transport provider, and it requires authorization by the responsible actor or a standing authorization under a reviewed protocol for that action type. Tier 3 covers external commitments and any action that touches clinical intent, such as booking an appointment, submitting an authorization request, or arranging transport, and each such action needs an individual human authorization, with clinical changes reserved for a clinically authorized actor. In Fig. 1 this gate sits on the path from burden comparison to the controlled executor, so a ready task of Tier 2 or 3 waits there for a human decision. In the worked example, the coverage and facility lookups run as soon as they are ready, while booking and transport wait for sign-off.

Worked Example of Diabetic Retinal Screening

A primary care clinician approves nonurgent diabetic retinal screening within 60 days for a synthetic patient who prefers Spanish, works overnight shifts, has no private vehicle, has limited portal access, needs coverage verification, and does not want unrelated diagnoses shared with a transport provider. A conventional entry would instruct staff to refer, arrange transportation, and remind the patient, leaving open which facility performs the test, who pays, who books transport, and what happens when no appointment fits. Fig. 2 traces how the compiler resolves these questions.

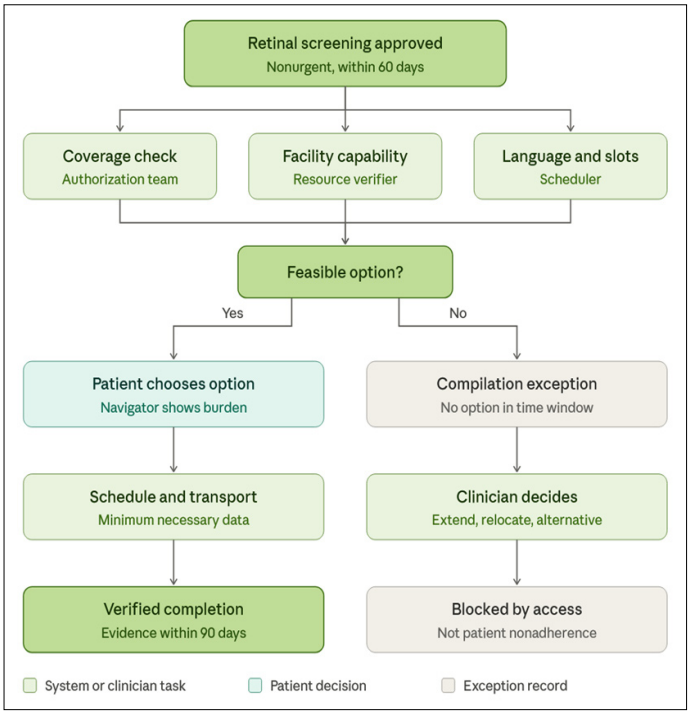


Fig. 2. Compilation of a diabetic retinal screening recommendation for a synthetic patient

The compiler opens three inquiry tasks with system-side owners for coverage, facility capability, and Spanish-language slots compatible with overnight work, and none of them asks the patient to search or call. When a feasible option exists, a navigator presents the options with their burden, the patient chooses, and transport is booked with the minimum necessary data. The episode closes when an encounter record or an authenticated result shows an examination performed within the 60-day window, and the evidence must arrive within 90 days of referral approval. When no option fits the window, the compiler returns an exception and leaves the order intact, and the clinician may extend the window, choose another site, or approve an alternative modality. Autonomous AI screening at the point of care is one such modality, and in a randomized trial among young people with diabetes, all participants in that arm completed the examination, compared with 22% of those referred to eye care (Wolf et al., 2024). If the clinician keeps the original order, the episode enters the record as blocked by access, the end state of the right branch in Fig. 2, and any other decision of the clinician returns the plan to compilation.

Staged Evaluation Program

The evaluation program tests six hypotheses: readiness checks reduce infeasible tasks labeled ready; intent protection reduces unauthorized substitutions; burden-aware planning yields options that patients judge more feasible; compiler assistance raises verified completion; serious errors stay at the level of usual care; and gaps between population groups grow no wider. Four stages test them, and each transition depends on prespecified safety thresholds. The first stage is participatory design with patients, caregivers, clinicians, navigators, interpreters, and privacy staff. The second stage is a benchmark of synthetic cases across referral, imaging, medication access, rehabilitation, and screening, with missing facts, expired authorizations, and events that change resources during an episode. Its comparators range from static instructions and general language models to ungated agentic planners and human navigators, and its primary metrics are the clinical intent preservation rate and the share of ready tasks that violate a hard constraint.

The third stage runs the compiler in silent mode on live data without patient contact and compares its decisions with navigator actions and observed availability before any progression (Vasey et al., 2022). Progression from this stage also requires the volume of inquiry tasks to stay within measured navigator capacity, because a system that opens inquiries faster than staff can resolve them adds burden. The fourth stage is a cluster randomized pilot in primary care, with screening completed within the clinical window and verified within 90 days of referral approval as the primary endpoint, and with completion at any time within 90 days, burden, navigator time, privacy events, and completion by language, disability, insurance, and digital access as secondary outcomes. A generalized linear mixed model with clinic-level clustering estimates the effect. Figures from the adjacent step of the pathway, where 35.1% of screen-positive patients reached a retina specialist within 1 year (Dahlan et al., 2023)

and navigation raised in-office examination attendance from 38.1% to 51.8% (Hark et al., 2025), inform the expected effect size, while local baseline data on screening completion determine the final sample size. The same figures set human navigation as the standard that the compiler must match.

DISCUSSION

The framework shifts responsibility for missed care. Current records mix six questions: what care was approved, what work it requires, who owns that work, whether the work is feasible, what burden it imposes, and whether care was delivered. The compiler keeps them apart and records each failure at its layer, so a missed screening becomes a statement about coverage, capacity, ownership, or patient choice. The rule that an unknown fact cannot support a ready task carries the greatest practical value, because it prevents a plausible plan from passing as confirmed.

The framework sits on top of existing standards, with FHIR CarePlan or ServiceRequest supplying the protected intent (HL7 International, 2023) and Task, Appointment, Consent, and Provenance resources carrying the compiled work. Its relation to navigation runs deeper, since navigators already perform most of the compilation steps and achieve completion in trials (Hark et al., 2025; Nelson et al., 2025). The compiler gives that tacit work named owners, recorded evidence, and audit trails, and it keeps navigators in review for cases that call for judgment. The compiler metaphor holds element by element: the recommendation as source code, the task graph as an intermediate representation, readiness checks as type checking, and exceptions as compile errors; it ends where human preference, clinical judgment, and scarcity resist formalization. The metaphor comes from the tradition of computer-interpretable guidelines, where PROforma models clinical knowledge as plans, decisions, actions, and enquiries (Sutton & Fox, 2003), GLIF3 maps guideline logic onto patient data for execution (Boxwala et al., 2004), and Peleg (2013) reviews the methods of the field. The Care Plan Compiler inherits that discipline and changes its object, since it takes the clinical recommendation as fixed input and compiles the operational layer beneath it.

Implementation depends on the facts that the readiness checker can obtain in machine-readable form, and their availability differs by fact class. The protected specification and the task graph map onto FHIR CarePlan, PlanDefinition, ServiceRequest, Task, Provenance, and AuditEvent resources, which are mature (HL7 International, 2023). Patient-confirmed social context can follow the Gravity Project implementation guide, which covers screening, health concerns, goals, and referrals for social needs (HL7 International, n.d.-c). Coverage requirements and prior authorization have dedicated Da Vinci guides for payer and provider exchange (HL7 International, n.d.-a, n.d.-b), and in the United States the CMS Interoperability and Prior Authorization final rule requires impacted payers to operate a prior authorization API from January 1, 2027 (Centers for Medicare & Medicaid Services, 2024). Other facts lack a comparable source. Open appointment slots stay inside facility scheduling systems in most settings, interpreter availability has no standard, the physical and sensory accessibility of a facility has no machine-readable representation, and provider directories exist but go stale. An early deployment will therefore resolve many tasks to inquiry required. That outcome has value of its own, because each inquiry becomes an explicit, owned, time-stamped, and countable task, which turns the tacit work of navigators into a measurable workload and produces the operational data that later automation would need. The configuration described here is specific to the United States, while the readiness logic, the task states, the consequence tiers, and the recompilation rule carry over to other health systems, where the binding facts may be facility capacity, medicine stock, distance, or the availability of community health workers.

Safety follows from permissions. Consequence tiers extend this logic, since an internal lookup passes on deterministic checks, while an action that contacts a person, discloses data, or commits the organization waits for a human decision whatever confidence the planner reports. The planner holds no credentials, the executor accepts only ready tasks, and retrieved content counts as data with no authority over policy, so human oversight resides in the system's access model. Equity analysis follows this pathway because a group's completion rate equals the product of its transition probabilities, and groups with less reliable records face a higher risk that inferred context will mislead the planner (Zink et al., 2026). Lower patient burden may raise navigator workload, so the evaluation tracks where work moves, and no-feasible-option results expose scarcity that the compiler cannot remedy.

The framework has limitations. Firstly, the search was targeted, and overlapping work may exist in unindexed or proprietary sources. Secondly, the represented constraints in the readiness properties depend on a correct implementation, and the worked example rests on a synthetic patient. Thirdly, part of the burden is subjective and changes over time. Fourthly, scarcity may outweigh any planning gain. Fifthly, the first pilot covers retinal screening alone, and legal rules for automated action differ across jurisdictions.

CONCLUSION

A clinical recommendation can be appropriate and still impossible to carry out. The Care Plan Compiler treats clinician-approved intent as a protected specification, compiles it into tasks with named owners, releases each task through a deterministic readiness check, compares feasible options by patient burden, and recompiles affected tasks when circumstances change. Its exception codes record each missed service at the layer where the system failed, and its staged evaluation gives the framework a defined path

to clinical testing. Under this design, a missed screening stops counting against the patient by default and becomes a measurable property of the care system.

REFERENCES

1. Abdul Salam, A. (2026). *Care Compiler* [Computer software]. Devpost. https://devpost.com/software/care-compiler
2. Bedi, S., Liu, Y., Orr-Ewing, L., Dash, D., Koyejo, S., Callahan, A., Fries, J. A., Wornow, M., Swaminathan, A., Lehmann, L. S., Hong, H. J., Kashyap, M., Chaurasia, A. R., Shah, N. R., Singh, K., Tazbaz, T., Milstein, A., Pfeffer, M. A., & Shah, N. H. (2025). Testing and evaluation of health care applications of large language models: A systematic review. *JAMA, 333*(4), 319–328. https://doi.org/10.1001/jama.2024.21700
3. Boxwala, A. A., Peleg, M., Tu, S., Ogunyemi, O., Zeng, Q. T., Wang, D., Patel, V. L., Greenes, R. A., & Shortliffe, E. H. (2004). GLIF3: A representation format for sharable computer-interpretable clinical practice guidelines. *Journal of Biomedical Informatics, 37*(3), 147–161. https://doi.org/10.1016/j.jbi.2004.04.002
4. Centers for Medicare & Medicaid Services. (2024). Advancing interoperability and improving prior authorization processes: CMS Interoperability and Prior Authorization final rule, CMS-0057-F. *Federal Register, 89*, 8758. https://www.federalregister.gov/d/2024-00895
5. Dahlan, K., Suman, P., Rubaltelli, D., Shrivastava, A., Chuck, R., & Mian, U. (2023). In a large healthcare system in the Bronx, teleretinal triaging was found to increase screening and healthcare access for an underserved population with a high incidence of T2DM and retinopathy. *International Journal of Environmental Research and Public Health, 20*(7), Article 5349. https://doi.org/10.3390/ijerph20075349
6. Hark, L. A., Gorroochurn, P., Pizzi, L. T., Jutkowitz, E., Goulak, A. M., Maruri, S. C., Harizman, N., Horowitz, J. D., Park, L., Wang, Q., Diamond, D. F., Valenzuela, I. A., Torres, D. R., Wang, Y., Sharma, T., Liebmann, J. M., & Cioffi, G. A. (2025). Patient navigators improve in-office eye exam adherence after community eye screenings in a randomized clinical trial: NYC-SIGHT study. *American Journal of Ophthalmology, 274*, 54–66. https://doi.org/10.1016/j.ajo.2025.02.030
7. Hill, J. C., Njeru, J. W., Wieland, M. L., Rutten, L. J., Jacobson, D. J., Jenkins, G. D., & Fan, C. (2022). Disparities among limited English proficient patients in colon cancer screening with multi-target DNA stool test. *Journal of Immigrant and Minority Health, 24*(2), 556–559. https://doi.org/10.1007/s10903-021-01214-7
8. HL7 International. (n.d.-a). *Da Vinci Coverage Requirements Discovery FHIR implementation guide* (Version 2.2.1). Retrieved October 12, 2026, from https://hl7.org/fhir/us/davinci-crd/
9. HL7 International. (n.d.-b). *Da Vinci Prior Authorization Support FHIR implementation guide* (Version 2.2.1). Retrieved October 12, 2026, from https://hl7.org/fhir/us/davinci-pas/
10. HL7 International. (n.d.-c). *SDOH Clinical Care for Multiple Domains: Gravity Project FHIR implementation guide* (Version 3.0.0). Retrieved October 12, 2026, from https://hl7.org/fhir/us/sdoh-clinicalcare/
11. HL7 International. (2023). *HL7 FHIR Release 5* (Version 5.0.0). https://hl7.org/fhir/R5/
12. Kyle, M. A., & Frakt, A. B. (2021). Patient administrative burden in the US health care system. *Health Services Research, 56*(5), 755–765. https://doi.org/10.1111/1475-6773.13861
13. Medda, F., & Gong, H. (2026). *Governed AI-agent coordination for dementia care: Architecture, safety contracts, and evidence-derived workflow verification* (arXiv:2609.25956). arXiv. https://doi.org/10.48550/arXiv.2609.25956
14. Nelson, H. D., Cantor, A. G., Pappas, M., Blackie, K., Yu, Y., & Fu, R. (2025). Patient navigation services for breast and cervical cancer screening and follow-up: A meta-analysis. *JAMA Internal Medicine, 185*(8), 976–985. https://doi.org/10.1001/jamainternmed.2025.1590
15. Njei, B., Al-Ajlouni, Y. A., Kanmounye, U. S., Boateng, S., Nguefang, G. L., Njei, N., Hamouri, S., & Al-Ajlouni, A. F. (2026). Artificial intelligence agents in healthcare research: A scoping review. *PLOS ONE, 21*(2), Article e0342182. https://doi.org/10.1371/journal.pone.0342182
16. Peleg, M. (2013). Computer-interpretable clinical guidelines: A methodological review. *Journal of Biomedical Informatics, 46*(4), 744–763. https://doi.org/10.1016/j.jbi.2013.06.009
17. Sutton, D. R., & Fox, J. (2003). The syntax and semantics of the PROforma guideline modeling language. *Journal of the American Medical Informatics Association, 10*(5), 433–443. https://doi.org/10.1197/jamia.M1264
18. Vasey, B., Nagendran, M., Campbell, B., Clifton, D. A., Collins, G. S., Denaxas, S., Denniston, A. K., Faes, L., Geerts, B., Ibrahim, M., Liu, X., Mateen, B. A., Mathur, P., McCradden, M. D., Morgan, L., Ordish, J., Rogers, C., Saria, S., Ting, D. S. W., . . . DECIDE-AI Expert Group. (2022). Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *Nature Medicine, 28*(5), 924–933. https://doi.org/10.1038/s41591-022-01772-9

19. Wolf, R. M., Channa, R., Liu, T. Y. A., Zehra, A., Bromberger, L., Patel, D., Ananthakrishnan, A., Brown, E. A., Prichett, L., Lehmann, H. P., & Abramoff, M. D. (2024). Autonomous artificial intelligence increases screening and follow-up for diabetic retinopathy in youth: The ACCESS randomized control trial. *Nature Communications, 15*, Article 421. https://doi.org/10.1038/s41467-023-44676-z

20. World Health Organization. (n.d.). *SMART guidelines*. Retrieved October 10, 2026, from https://www.who.int/teams/digital-health-and-innovation/smart-guidelines

21. Zink, A., Luan, H., & Chen, I. Y. (2026). Access to care affects electronic health record reliability and AI-driven disease prediction. *Nature Health, 1*(3), 316–325. https://doi.org/10.1038/s44360-026-00054-9

Copyright: © 2026 The Author(s). This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.