GMR
The Original

AUTONOMOUS MULTI-AGENT AI FRAMEWORK FOR END-TO-END CLINICAL DECISION SUPPORT AND CARE COORDINATION

Bharath Penaka[1]

[1]Independent Researcher, bharathpenakaa@gmail.com

ABSTRACT—Current clinical decision support (CDS) systems are usually integrated into electronic health record (EHR) applications and primarily focus on clinical diagnostics and treatment recommendations. In contrast, the developed autonomous multi-agent AI framework extends CDS capabilities to all aspects of patient candidate care across the entire care continuum. From a first-principles perspective, multi-agent systems (MASs) are defined, and the key applications of MASs in healthcare are outlined. CDS systems support every clinical decision that frontline clinicians must make during health system interactions, yet the sheer number of decisions constitutes an insurmountable barrier for human intelligence alone. Extensive decision support for every frontline interaction in a health system is provided by an autonomous MAS with the stated capabilities in a principled objective manner. Autonomous MASs represent a new paradigm for end-to-end CDS. The canonical two- to five-way integration by clinical service line forms the theoretical foundation linking application breadth and the need for such an autonomous end-to-end MAS. Autonomy extends efforts in centralized and multi-level support and achieves an entirely new level of CDS candidacy, covering every interaction with a health system throughout the health continuum. As an end-to-end solution utilizing every available EHR data source across the full data richness span, it closes the big data, explainability, and big data yet little intelligence per interaction chunks associated with complimentary integration. The design addresses safety, reliability, explainability, and trust—all elements of the human operating picture. Validation, while a near impossibility in another context, is the master fulcrum of root-cause evidence-based maturity, independent of outcome and thus spare to the routine.

KEYWORDS—Autonomous Multi-Agent AI, Clinical Decision Support, Care Coordination, Multi-Agent Systems, Agentic AI, Healthcare Artificial Intelligence, Clinical Workflow Automation, Electronic Health Records, Personalized Patient Care, Intelligent Healthcare Systems.

1. INTRODUCTION

Artificial Intelligence (AI) holds the promise of revolutionising global healthcare delivery by enhancing the quality and accessibility of clinical decision support (CDS) systems for diagnosis, risk stratification and prognostication. AI researchers are, therefore, developing increasingly sophisticated and capable learning algorithms. However, in their practical implementation, AI algorithms remain individualistic, unsupervised and unmonitored. That is, they function independently of every other agent in the hospital and healthcare ecosystems, other than sharing task-specific information through poorly-defined interfaces and interaction protocols. As a result, AI algorithms are best regarded as information engines that contribute to the broader CDS architecture.

Effective healthcare delivery involves the coordinated interactions of a myriad of stakeholders, both clinical and non-clinical, working together towards a common goal of providing high-quality care to patients. Each stakeholder has a unique set of capabilities, and an associated set of information needs. Information engines develop AI algorithms that offer assistance in specific task areas. Current approaches to integrating AI into hospital operations primarily consider the perspective of clinical delivery; that is, the delivery of care to patients by any clinical care team. While the outcome of such a process is important, the focus so far ignores the fact that clinical care is not the only activity that is ongoing in a hospital at any time. Internal operations within the hospital and the broader healthcare ecosystem need to keep operating for the hospital to be able to manage and support the delivery of high-quality clinical care to patients. Consequently, the proposed multi-agent AI framework does not seek to interfere directly with clinical delivery, but rather mine the diverse data sources available in real-time to provide supervisory support to every other stakeholder involved in providing clinical delivery at the hospital, as well as support of the overall clinical delivery ecosystem.

Table 1. Core Stakeholder Agents and Roles

Agent Type	Represents	Core Function	Interaction Mode
Clinical Agent	Clinicians	Diagnostic / treatment	Leads during FCW*

		recommendation support	
Nursing / Allied Health Agent	Nursing & allied staff	Care-process execution and monitoring	Collaborates during FCW
Care Coordination Agent	Patients	End-to-end journey management across the continuum	Continuous (before/during/after FCW)
Administrative Agent	Hospital operations	Resource allocation, scheduling, capacity	Negotiation-based
Payer / Compliance Agent	Payers / regulators	Reimbursement checks, regulatory adherence	Asynchronous
Specialty / Consult Agent	Specialists & committees	Planning and reasoning support for referrals	On-demand

2. Background and Motivation

Most clinical decision-support and care-coordination solutions are siloed, tightly coupled, and ineffective for complex cases. A fully autonomous, collaborative multi-agent-as-a-service platform, supporting stakeholder needs throughout the disease continuum, is proposed. It facilitates expert negotiation, mediates conflicts, mitigates misunderstandings, generates explainable agent decisions, operates in partnership with humans, and complies with regulations. The approach opens possibilities for distributed, cross-institution learning and training across healthcare. Clinical Decision Support (CDS) aims to improve clinical performance and patient outcomes. At its most advanced, it provides intelligent recommendations during clinical encounters, complementing a clinician's knowledge and experience. Care Coordination (CC) enables the efficient management of complex clinical journeys. Such CDSS/CDS services are delivered as tightly coupled, siloed tools from single vendors, focusing on well-defined operations or diseases. A fully autonomous, end-to-end multi-agent platform supporting multiple stakeholders across the disease continuum would be a game changer. Such Multi-Agent Systems (MAS) would facilitate collaboration, negotiation, conflict mediation, and successful interaction with humans (via Explainable Artificial Intelligence/ XAI). Moreover, a solution encompassing all ICAARE elements of healthcare—Intelligent, Commumunity-Driven, Adaptive, Accessible, Resilient, Equity-Minded—and capable of learning from multi-institution activity would be deeply transformative.

Table 2. Functional vs. Nonfunctional Requirements

Category	Requirement	Design Mechanism
Functional	Support every clinical encounter	Continuous agent coverage across all FCW touchpoints
Functional	Auditability	Immutable audit trail of agent actions & decisions
Nonfunctional	Safety	Formal verification of safety properties per agent
Nonfunctional	Reliability	Redundant supervision by appropriate authorities
Nonfunctional	Resilience	Graceful operation in normal, atypical, emergency modes

Nonfunctional	Explainability	XAI-based rationale when suggestions deviate from norm
Nonfunctional	Trust / Governance	Human oversight, veto authority, regulatory compliance

3. Conceptual Framework: Autonomous Multi-Agent Systems in Healthcare

A multi-agent system consists of multiple autonomous agents acting in concert to achieve goals. Each agent has a unique identity and its own set of objectives, capabilities, knowledge, and constraints. Two essential features that characterize a multi-agent system are its use of software agents and its capability for intelligent interaction and cooperation among agents. An autonomous multi-agent system integrates intelligent devices capable of sensing, thinking, and responding to environmental changes, and use complex levels of coordination to deliver adaptive care. Autonomous multi-agent systems (AMAS)—defined as self-organizing solutions to problems too complex for an individual, monolithic system—provide natural, proven support for all aspects of clinical decision support and care coordination. Each stakeholder actor is represented by a set of agents that acts on its behalf and satisfies its needs. Stakeholder support is natively implemented through negotiation and conflict resolution, enabling risks to be mitigated without a monolithic authority. Interoperability across agents is facilitated through standards-based communication, allowing integration with existing systems. Using performance-execution formalism, these properties meet the traditional, safety-critical requirements for intelligent systems: well-defined behavior, explainability, trustworthiness, reliability, and safety. Requirements resulting from clinical workflows mandated by professional societies; data access mandated by laws, regulation, and reimbursement; and clinical supervision of all actions and care coordinators.

Table 3. Data Sources in the Perception & Integration Layer

Source	Data Type	Primary Processing Step
EHR	Structured & unstructured records	Normalization, terminology / ontology mapping
Imaging	DICOM / pixel data	Quality certification, provenance tracking
Laboratory	Structured results	Heterogeneity reduction across systems
Genomics	Sequence / variant data	Cohort-defining curation, FDA-cleared pipelines
Wearables / IoT	Streaming biosignals	Real-time noise filtering, context-aware alerting
External / SDOH	Social determinants, literature	Semantic harmonization into the Knowledge Graph

A. Agents, Roles, and Interactions

Conceptualizing a Multi-Agent AI Framework for CDSS and Care Coordination Stakeholders in the clinical setting include patients, clinicians, nursing and allied health staff, specialty care, administrative services, payers, and healthcare organizations. Multi-agent systems differentiate individual entities, represent their motivations, and model capability gaps. Agent-based clinical decision support and care coordination therefore encompass diverse components beyond diagnosis and recommendation. These include real-time representations of patient status with respect to care processes, the capacity to coordinate care with, or on behalf of, the real-time clinical team, and fulfilling stakeholder requirements using patient-centric approaches both before and during formal clinical workflows. Successful clinical operations depend on a combination of these capabilities. To fulfill the needs of patients and clinical teams, agents must work collaboratively with the real-time clinical team, including the nursing and allied health staff. Clinical workflows employ a combination of plans, hints, and descriptions for knowledge representation. Team members communicate intent via discussion or formalized mutual plans and indicate issues either verbally or through non-verbal hints. Failure detection, mutual negotiation, alternative initialization, and continuation or cessation are enabled by social behaviours, business rules, and resolution strategies. For negotiation with management or administrative teams, ability to describe state and events is critical. Support for planning and reasoning enhances usability for specialty services or committees. For maintenance teams, real-time-based intent is important.

These diverse application roles include not only the traditional functions of clinical decision support with interactive hints but also those of care coordination before, during, and after Formal

Clinical Workflows (FCW). Core components designed to fulfill specific roles must collaborate, exchanging knowledge and intent to enable the overall multi-agent system capability. Collaboration occurs in two modes: for real-time clinical operations during FCW and for any other function at any other time. During FCW, true clinical agents must lead, fulfilling formal needs and employing their enriched real-time knowledge of patient stability and process compliance. At all other times, the broader system maintains overall social behavior while providing enriched assistance for real-time clinical agents.

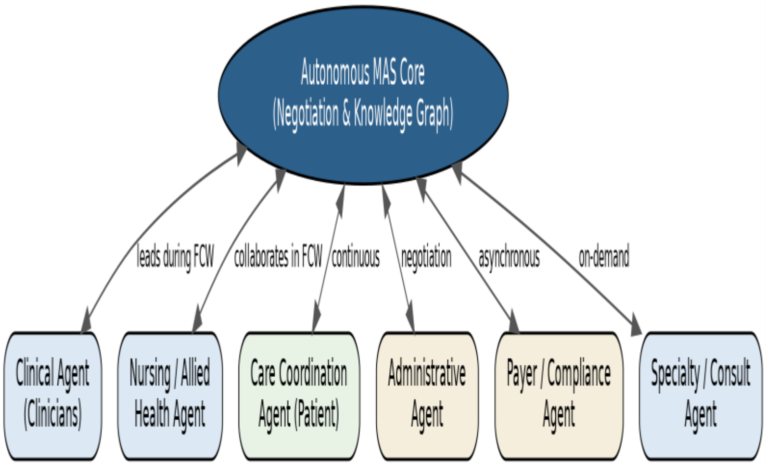

Fig. 1. Each stakeholder is represented by a dedicated agent that negotiates through the shared MAS core / knowledge graph, consistent with the collaboration modes described in Section 3.A.

4. System Requirements and Design Principles

Functional and nonfunctional requirements for the multi-agent autonomous decision support and coordination system are specified. Design principles address safety, reliability, and resilience; trust, explainability, and governance; compliance with clinical workflows; and adherence to pertinent security, auditability, regulatory, and data-sharing requirements. Nonfunctional requirements, informed by the clinical decision support and care coordination goals articulated, concern safety, reliability, and resilience. Safety is paramount: the system must not expose patients or healthcare professionals to unacceptable risk. It should also be reliable, providing major support for and coordination during every patient encounter, and resilient, enabling operation in normal, atypical, and emergency conditions. Mechanisms are therefore proposed to meet the requirements: all agents undergo formal verification of safety properties; and supervision by appropriate authorities, together with an explicit explanation of why suggestions deviate from expectations when identified, ensure that users can spot errors of fact and deviations from the norm. Functional, nonfunctional, and safety requirements capture the need for support during all clinical encounters, an audit trail demonstrating system integrity, and compliance with regulations incorporating these desiderata. Other desiderata address the elucidation of system processes using appropriate normal-or-abnormal intentional-agent models, the reduction of sensitive information-sharing requirements, and the systematic propagation of error degradation through the system in order to support the early identification of emerging errors.

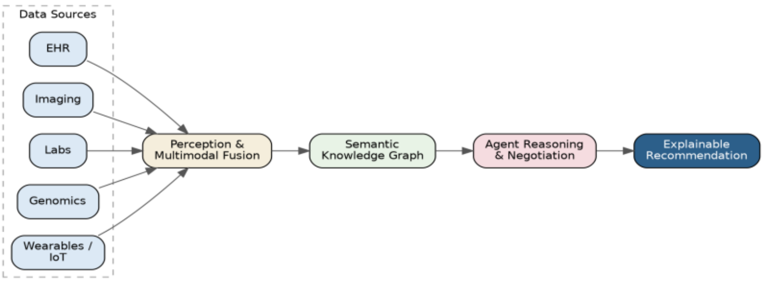

Fig. 2. Multimodal data sources are normalized and fused, integrated into a semantically harmonized knowledge graph, and consumed by agent reasoning to produce explainable recommendations (Section 5.A).

A. Safety, Reliability, and Trustworthiness

Autonomous systems governing human affairs must meet safety, reliability, and resilience criteria appropriate to their domain. As the consequences of failure can be severe, trusted system design is fundamental. In healthcare, oversight is essential to ensure that clinical workflows are respected and that regulatory compliance is observable. This subsection proposes safety and regulatory considerations for the multi-agent framework.

Safety and reliability criteria must accommodate the very different domains of child protection and healthcare, aspects such as whose interests and what aspects of safety are paramount, the need for whatever implicit trust underlies human caring and workplaces, and the possibility of negation or veto rather than action. Specific trust, explainability, and governance requirements must be defined for each domain.

The architecture is functionally modular, facilitating independently designed and tested components. Safety, reliability, and trustworthiness are guaranteed by specifying external interfaces, data communication standards, security requirements, audit trails for actions and decisions, respect for clinical workflow affordances, and compliance with relevant regulatory frameworks and norms. Safety and explainability requirements complement Endorsed Security Level specifications in supporting Trust Services in a cyber-risk context.

5. Agent Modeling and Reasoning

Automated agent-based systems must assemble and evaluate comprehensive multimodal knowledge to adequately describe health conditions, predict disease trajectories, classify treatment outcomes, and devise individualized management strategies. Design, implementation, and integration of these knowledge-intensive components for major stakeholder agents are elaborated. An autonomous multi-agent system for clinical decision support during critical illness management serves as a motivating example.

Comprehensive patient-specific descriptions require contextualized fusion of information from diverse data sources. Major sources include electronic health records, imaging, laboratory tests, genomics, and data from wearable sensors for biosignal monitoring. Specialized pipelines normalize information generated by distinct systems, ensuring interoperability. Supporting ontologies codify common medical entities and concepts, and pipeline operation conducts routine assessments of terminology compatibility. Critical evaluation of data provenance at fusion time strengthens knowledge reliability.

The Agent Models and Reasoning section elaborates the development of knowledge-intensive core components for each of the major stakeholder agents. These core components encompass multimodal health condition representation, individual disease trajectory forecasting, model-driven outcome classification, and agent-based elaboration of personalized treatment plans. Combined component capabilities effectively simulate the human-staffed decision-making process.

Mathematical Formulas:

$$MAS = \langle A, E, I, G \rangle \tag{1}$$

Eq. (1): a multi-agent system is the tuple of agents $A = \{a_1, \ldots, a_n\}$, environment E (the hospital/EHR data ecosystem), interaction protocol I, and shared goal G (patient-centered care quality).

$$U_i(s) = w_1 \cdot Safety_i(s) + w_2 \cdot Reliability_i(s) + w_3 \cdot Explainability_i(s) - w_4 \cdot Risk_i(s) \tag{2}$$

Eq. (2): the utility of agent i for state s combines the safety, reliability, and explainability criteria of Section 4, penalized by residual risk; w_k are stakeholder-specific weights.

$$F(x) = \Sigma_k\ \alpha_k \cdot \varphi_k(x_k), \quad \Sigma_k\ \alpha_k = 1 \tag{3}$$

Eq. (3): multimodal perception fusion, where x_k is data from modality k (EHR, imaging, labs, genomics, wearables), φ_k a per-modality encoder, and α_k a learned fusion weight (Section 5.A).

$$d^* = \text{argmax\_}(d \in D)\ \ \Sigma_i\ \lambda_i \cdot U_i(d) \quad \text{subject to} \ \ Safety(d) \geq \theta \tag{4}$$

Eq. (4): negotiated consensus decision d* maximizes weighted stakeholder utility subject to a minimum safety threshold θ, formalizing the negotiation and conflict-resolution behavior described in Section 3.

A. Perception and Data Integration

Data sources include structured information from electronic health records (EHR), imaging studies, laboratory analyses, and genomic tests, as well as unstructured data, such as clinical notes or reports. Perception entails the preparation, purification, and organization of data for downstream analysis. The relevant data collected by a healthcare organization are synthesized into a data lake with well-defined pipelines for heterogeneity reduction, normalization, and quality certification. Provenance tracking includes the origin, chain of transformations, and temporal footprint of specific measures. Sensors, wearables, and portable health devices provide processed and integrated data streams as Internet of Things (IoT) sources. Temporal precision and noise levels can thus be relaxed, and complementary information combined to generate context-aware alerts. Imaging and genomic facilities supply data of varying quality, up-to-date chain-of-custody information, and additional volumes for cohort-defining projects. Large-scale datasets correspond to well-curated FDA-cleared supervised-learning pipelines. These data are integrated along with social determinants from external sources and recent publications for a semantically harmonized Knowledge Graph. The performance of multisensory data integration has a direct effect on the operation and health of the end user. Automated and autonomous data fusion of multiple modalities should play a crucial role—identifying interventions, mitigating conflicting inputs, or selecting the best information stream for specific tasks. State-of-the-art methods for early detection of health deteriorations that exploit unbalanced high-dimensional label-deficient data can be employed for real-time quality check of IoT alerts, as well as for fusion of all monitored data and for relations with other agents. Perception is also involved in any task in which a high-dimensional signal must be processed or analyzed.

6. CONCLUSION

An autonomous, multi-agent system architecture is proposed to address clinical decision support and care coordination problems. Relevant stakeholders are modeled as distinct agents whose responsibilities are based on decision- and care-coordination needs. The needs mapping highlights functional gaps that need to be bridged by the agents, including collaborative decision support, coordination of multiple sources of evidence, negotiation and conflict resolution, engagement with clinical workflows, and health risk analysis. To support these roles, the agents are endowed with capabilities for multi-party engagement, intelligent negotiation, bidirectional traffic accident prevention, medical image processing, risk analysis, and fusion of complex multimodal evidence. These capabilities enable seamless communication and collaboration among the agents in a variety of configurations and processes. The proposed end-to-end architecture is driven by workflow safety and reliability requirements derived from clinical consensus guidelines. The safety principles also offer mechanisms for establishing the trustworthiness of non-clinical agent processes and decisions. The architecture is designed to meet both functional and non-functional requirements and to uphold standards for auditability, security, and GDPR compliance. Consistency and accuracy of health records and other evidence sources are explicitly integrated into the architecture, aligning supply readiness with demand and supporting reliable decision-making and care. The framework draws on established work on autonomous multi-agent systems while meeting the specific requirements of the healthcare domain. Once developed, agents representing clinical stakeholders will enable end-to-end prevention, detection, and management of traffic accidents in the Australian healthcare system.

REFERENCES

[1] Klang, E., Omar, M., Raut, G., Agbareia, R., Timsina, P., Freeman, R., Gavin, N., Stump, L., Charney, A. W., Glicksberg, B. S., & Nadkarni, G. N. (2026). Orchestrated multi agents sustain accuracy under clinical-scale workloads compared to a single agent. npj Health Systems, 3, 23.

[2] Liu, P., & Xiao, L. (2026). An intention-oriented multi-agent dialogue system for patient-centered decision-making. BMC Medical Informatics and Decision Making, 26, 276.

[3] Hegselmann, S., von Arnim, G., Rheude, T., Kronenberg, N., Sontag, D., Hindricks, G., Eils, R., & Wild, B. (2026). Large language models are powerful electronic health record encoders. npj Digital Medicine, 9, 530.

[4] Kirchler, M., Ferro, M., Lorenzini, V., van de Water, R. P., FinnGen, Lippert, C., & Ganna, A. (2026). Large language models improve transferability of electronic health record-based predictions across countries and coding systems. npj Digital Medicine, 9, 177.

[5] Kolla, S. K., & Bandi, V. D. V. K. (2025). Autonomous Clinical Monitoring Platforms Using Reinforcement Learning and Deep Neural Networks. International Journal of Advanced Research in Computer Science & Technology (IJARCST), 8(6), 13300-13313.

[6] Nargesi, A. A., You, J. G., Bitterman, D. S., Succi, M. D., Mishuris, R. G., Poon, E. G., & Landman, A. B. (2026). Tracing the pen: Electronic health records amid the rise of generative AI. npj Digital Medicine, 9, 482.

[7] Zhuang, Y., Wang, B., Yin, C., et al. (2026). Clinical large language model centered on electronic medical records. npj Digital Medicine, 9, 511.

[8] Gaber, F., Shaik, M., Allega, F., Bilecz, A. J., Busch, F., Goon, K., Franke, V., & Akalin, A. (2025). Evaluating large language model workflows in clinical decision support for triage and referral and diagnosis. npj Digital Medicine, 8, 263.

[9] Li, D., Jiang, N., Huang, K., et al. (2025). Streamlining evidence based clinical recommendations with large language models. npj Digital Medicine, 8, 793.

[10] Li, J., et al. (2025). Enhancing diagnostic capability with multi-agents conversational large language models. npj Digital Medicine, 8, 159.

[11] Masanneck, L., Meuth, S. G., & Pawlitzki, M. (2025). Evaluating base and retrieval augmented LLMs with document or online support for evidence based neurology. npj Digital Medicine, 8, 137.

[12] Mangalampalli, B. M. (2024). AI-Enhanced Data Governance: Automating Compliance In Healthcare Analytics Platforms. The Review Of DIABETIC STUDIES OPEN ACCESS.

[13] Wiest, I. C., Bhat, M., Clusmann, J., Schneider, C. V., Jiang, X., & Kather, J. N. (2025). Large language models for clinical decision support in gastroenterology and hepatology. Nature Reviews Gastroenterology & Hepatology, 22(11), 773–787.

[14] Yang, R., Ning, Y., Keppo, E., Liu, M., Hong, C., Bitterman, D. S., Ong, J. C. L., Ting, D. S. W., & Liu, N. (2025). Retrieval-augmented generation for generative artificial intelligence in health care. npj Health Systems, 2, 2.

[15] Zhang, G., Xu, Z., Jin, Q., Chen, F., Fang, Y., Liu, Y., Rousseau, J. F., Xu, Z., Lu, Z., Weng, C., & Peng, Y. (2025). Leveraging long context in retrieval augmented language models for medical question answering. npj Digital Medicine, 8, 239.

[16] Dennstädt, F., Hastings, J., Putora, P. M., Schmerder, M., & Cihoric, N. (2025). Implementing large language models in healthcare while balancing control, collaboration, costs and security. npj Digital Medicine, 8, 143.

[17] Jonnagaddala, J., & Wong, Z. S. Y. (2025). Privacy preserving strategies for electronic health records in the era of large language models. npj Digital Medicine, 8, 34.

[18] Johri, S., Jeong, J., Tran, B. A., Schlessinger, D. I., Wongvibulsin, S., Barnes, L. A., Zhou, H. Y., Cai, Z. R., Van Allen, E. M., Kim, D., Daneshjou, R., & Rajpurkar, P. (2025). An evaluation framework for clinical use of large language models in patient interaction tasks. Nature Medicine, 31, 77–86.

[19] Omar, M., Nadkarni, G. N., Klang, E., & Glicksberg, B. S. (2024). Large language models in medicine: A review of current clinical trials across healthcare applications. PLOS Digital Health, 3(11), e0000662.

[20] Maguluri, K. K. (2025). Natural language processing in healthcare: unlocking insights from clinical data. How Artificial Intelligence is Transforming Healthcare IT: Applications in Diagnostics, Treatment Planning, and Patient Monitoring, 72.

[21] Hager, P., Jungmann, F., Holland, R., Bhagat, K., Hubrecht, I., Knauer, M., Vielhauer, J., Makowski, M., Braren, R., Kaissis, G., & Rueckert, D. (2024). Evaluation and mitigation of the limitations of large language models in clinical decision-making. Nature Medicine, 30, 2613–2622.
[22] Van Veen, D., Van Uden, C., Blankemeier, L., Delbrouck, J. B., Aali, A., Bluethgen, C., Pareek, A., Polacin, M., Reis, E. P., Seehofnerová, A., Rohatgi, N., Hosamani, P., Collins, W., Ahuja, N., Langlotz, C. P., Hom, J., Gatidis, S., Pauly, J., & Chaudhari, A. S. (2024). Adapted large language models can outperform medical experts in clinical text summarization. Nature Medicine, 30, 1134–1142.
[23] Haltaufderheide, J., & Ranisch, R. (2024). The ethics of ChatGPT in medicine and healthcare: A systematic review on large language models (LLMs). npj Digital Medicine, 7, 183.
[24] Riedemann, L., Labonne, M., & Gilbert, S. (2024). The path forward for large language models in medicine is open. npj Digital Medicine, 7, 339.
[25] Tam, T. Y. C., Sivarajkumar, S., Kapoor, S., Stolyar, A. V., Visweswaran, S., Mathur, P., Cacciamani, G. E., Sun, C., Peng, Y., & Wang, Y. (2024). A framework for human evaluation of large language models in healthcare derived from literature review. npj Digital Medicine, 7, 258.
[26] Kaczmarczyk, R., Wilhelm, T. I., Martin, R., & Roos, J. (2024). Evaluating multimodal AI in medical diagnostics. npj Digital Medicine, 7, 205.
[27] Singhal, K., Azizi, S., Tu, T., Mahdavi, S. S., Wei, J., Chung, H. W., Scales, N., Tanwani, A., Cole-Lewis, H., Pfohl, S., Payne, P., Seneviratne, M., Gamble, P., Kelly, C., Babiker, A., Schärli, N., Chowdhery, A., Mansfield, P., Demner-Fushman, D., Agüera y Arcas, B., et al. (2023). Large language models encode clinical knowledge. Nature, 620(7972), 172–180.
[28] Reddy, V. A. R. (2025). Journal of Rare Cardiovascular Diseases. Health, 5(3), 402-422.
[29] Moor, M., Banerjee, O., Abad, Z. S. H., Krumholz, H. M., Leskovec, J., Topol, E. J., & Rajpurkar, P. (2023). Foundation models for generalist medical artificial intelligence. Nature, 616(7956), 259–265.
[30] Clusmann, J., Kolbinger, F. R., Muti, H. S., Carrero, Z. I., Eckardt, J. N., Laleh, N. G., Löffler, C. M. L., Schwarzkopf, S. C., Unger, M., Veldhuizen, G. P., Wagner, S. J., & Kather, J. N. (2023). The future landscape of large language models in medicine. Communications Medicine, 3, 141.
[31] Lee, P., Bubeck, S., & Petro, J. (2023). Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. New England Journal of Medicine, 388(13), 1233–1239.
[32] Kanjee, Z., Crowe, B., & Rodman, A. (2023). Accuracy of a generative artificial intelligence model in a complex diagnostic challenge. JAMA, 330(1), 78–80.