Open Access
EDITORIAL
Journal of Biomedical Sciences
DOI: https://doi.org/10.3126/jbs.v12i2.91040

Artificial intelligence agents in medicine: Emerging capabilities and implementation challenges

Roy B

Corresponding author:
Dr. Bedanta Roy,
Associate Professor, Department of Physiology, Faculty of Medicine, Quest International University, Ipoh, Perak, Malaysia
Email: bedanta.roy@gmail.com ORCID

Information about the article:

Published online: Feb 18, 2026

Cite this article:

Roy B. Artificial intelligence agents in medicine: Emerging capabilities and implementation challenges. Journal of Biomedical Sciences. 2025;12(2):15-16

Publisher
Nepal Health Research Society, Bahundhara -6, Gokarnesowor Municipality, Kathmandu, Nepal
eISSN 2382-5545, ISSN 2676-1343 (Print)

Artificial intelligence (AI) has transformed many industries, and healthcare is no exception. Traditional AI systems include machine learning models that analyse medical imagery, predict clinical outcomes, and automate administrative workflows. In contrast, AI agents are designed to operate with greater autonomy—planning, adapting, and interacting with complex data or environments while executing goal-directed actions. Distinct from static algorithms or single-response conversational bots, these systems incorporate modules for iterative reasoning, memory, and the invocation of external tools, enabling personalised, interactive decisions tailored to specific clinical contexts.

The growth of large language models and multimodal architectures has accelerated the creation of intelligent agents that interpret clinical narratives, handle both structured and unstructured health data, and support dynamic problem-solving. However, the rapid advancement of technology often outpaces research on safety, efficacy, and ethical implementation in healthcare.

AI agents in healthcare generally fall into three main types: conversational agents that facilitate patient-clinician interactions, workflow or automation assistants that handle back-office and record-keeping tasks, and multimodal decision support systems that combine various data types for clinical decision-making. Conversational agents support patient engagement, symptom triage, or mental health interventions via natural language interfaces. Agents that apply cognitive-behavioural strategies have been explored in controlled studies of self-management and therapeutic support.

Workflow/Automation assistants demonstrate value in automating documentation, interpreting electronic health records, translating clinical queries, and extracting structured experimental conditions from biomedical datasets. Multimodal decision support combines text, imaging, and structured data to provide diagnostic insights and support planning complex treatments, such as radiotherapy optimisation. Additionally, multi-agent frameworks break down tasks into collaborative sub-agents, mirroring human interprofessional workflows.

Across archetypes, the core mechanisms include retrieval-augmented generation (RAG) for grounding outputs in real data, multi-agent debate loops for cross-validation, and self-debugging routines to refine outputs iteratively.

A significant limitation of existing research is the absence of validation in real-world clinical settings. Most AI evaluations are limited to simulated environments, retrospective data, or controlled laboratory experiments, and focus primarily on process-related outcomes, such as task efficiency or diagnostic accuracy. Conversely, few studies have examined patient outcomes, safety measures, or long-term performance in everyday clinical practice.

Theoretical research on agent architectures indicates that intelligent agents can assist clinician decision-making by emulating parts of human reasoning or by collaboratively verifying diagnoses and treatment plans. Nonetheless, these systems necessitate thorough prospective testing to confirm their reliability, usability, and clinical value beyond human standards. Although AI agents have great potential, safety remains a vital issue. Recent studies show that AI tools, especially large language models, can spread incorrect medical information if trained or prompted improperly, risking diagnostic or treatment mistakes in medical settings. These results highlight the importance of implementing strong safeguards, interpretability features, and harm-reduction design principles that maintain human oversight and accountability.

Ethical deployment also requires focus on fairness. Biases in training data can worsen disparities in diagnosis and treatment if not regularly checked and corrected. New advocacy efforts advocate for "equity-first" standards to steer AI development and ensure inclusive health outcomes. From a regulatory perspective, many healthcare AI systems have advanced faster than current oversight methods, complicating approval, post-market monitoring, and liability issues. Uncertainty persists over who should be responsible for AI errors, underscoring the need for clear governance.
AI agents represent a promising evolution in healthcare technology, offering capabilities beyond conventional AI systems through autonomous reasoning, adaptability, and multimodal integration. Innovation in conversational interfaces, workflow automation, and decision support remains confined to early settings with limited clinical validation. To transition from potential to practice, concerted efforts in evaluation, governance, and ethical deployment are needed. With structured oversight and interdisciplinary collaboration, AI agents can augment healthcare delivery while maintaining patient safety and clinician empowerment.

Dr. Bedanta Roy
Editor-in-Chief
Journal of Biomedical Sciences
18.2.2026

Keywords
Ethical, diagnosis, healthcare, legal, medical, social

Abbreviations
Artificial intelligence (AI), retrieval-augmented generation (RAG)

Availability of data and materials
Not applicable.

Competing interests
None declared.

Publisher's Note
NHRS remains neutral regarding jurisdictional claims in published maps and institutional affiliations.

The publisher shall not be legally responsible for any types of loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Author information
Dr. Bedanta Roy, Associate Professor, Department of Physiology, Faculty of Medicine, Quest International University, Ipoh, Perak, Malaysia ORCID