Iranian Journal of Blood & Cancer

Iran J Blood Cancer, 2026, 18(2), 107-120
https://ijbc.ir

Original Article

From Data to Decision: A Conceptual Architecture of Agentic AI for Hemato-Oncology Tumor Boards

Samira Babaei[1], Mustafa Ghaderzadeh[2*], Cirruse Salehnasab[3*]

[1]Department of Nursing, Naghadeh School of Nursing, Urmia University of Medical Sciences, Urmia, Iran.
[2]Boukan Faculty of Medical Sciences, Urmia University of Medical Sciences, Urmia, Iran.
[3]Social Determinants of Health Research Center, Yasuj University of Medical Sciences, Yasuj, Iran.

Scan and read the article online

Citation Babaei S, Ghaderzadeh M, Salehnasab C. From Data to Decision: A Conceptual Architecture of Agentic AI for Hemato-Oncology Tumor Boards. Iran J Blood Cancer. 2026 Jun 30;18(2): 107-120.

doi 0

Article info:
Received: 07 May 2026
Accepted: 20 Jun 2026
Published: 30 Jun 2026

Abstract

Background and Objective: Hematological malignancies are facing increasing complexity in diagnosis and treatment, which requires multidisciplinary decision-making in tumor boards. Traditional predictive AI models have limited clinical application due to limitations such as black-box nature and inability to reason in multiple steps. This study aimed to design a comprehensive conceptual framework for a multi-agent Agentic AI system architecture to support decision-making in hemato-oncology tumor boards.

Methods: This conceptual study was conducted using the Design Science Research methodological framework. The proposed architecture was developed based on the convergence of three reference frameworks, including the ReAct reasoning paradigm, the HL7 FHIR R4 standard, and the NIST AI RMF 1.1 framework. The system architecture was described using Unified Modeling Language (UML) 2.5, and four specialized agents (Data Parser, Genomics, Guideline RAG, and Orchestrator) were specified based on a standard seven-component design pattern. The architecture was evaluated using the multidimensional RAMSES framework by an eight-member multidisciplinary expert panel.

Results: The proposed architecture was designed with five distinct layers (data acquisition, agent coordination, knowledge and tools, safety and governance, and presentation). The system employed a multilayered Defense-in-Depth security approach with four independent protective layers, which reduced the hallucination rate from 18.4% to 3.2%. An application scenario based on an AML case showed that the system was able to process a complex clinical file in 3 minutes and 21 seconds and produce a 2,847-word structured report with a confidence score of 0.82. The RAMSES evaluation yielded an overall average score of 4.4 out of 5, across the three functional (4.7), technical (4.7), and clinical (4.3) dimensions.

Conclusion: This study provides a roadmap for the transition from predictive AI to agentic AI in hemato-oncology. The proposed multi-agent architecture, while preserving the human-in-the-loop principle, ensures reliability, explainability, and regulatory compliance, and can reduce the cognitive burden on tumor boards while democratizing access to precision medicine.

Keywords:
Artificial intelligence
Multi-agent system
Hemato-oncology
Tumor board
Clinical decision support
System architecture

* Corresponding Author:
Mustafa Ghaderzadeh
E-mail: Mustafa.ghaderzadeh@gmail.com

Cirruse Salehnasab
E-mail: cirruse.salehnasab@gmail.com

[Downloaded from ijbc.ir on 2026-11-13]

1. INTRODUCTION

Hematologic malignancies have undergone a fundamental transformation in their diagnostic and therapeutic paradigms over the past decade. The World Health Organization's fifth edition classification of tumors of the hematopoietic and lymphoid tissues (WHO-HAEM5) and the International Consensus Classification (ICC 2022), by integrating multiple layers of molecular, genetic, and immunophenotypic data alongside traditional morphology, have imposed a new level of complexity on the diagnostic process [1, 2]. For instance, in acute myeloid leukemia (AML), the identification of more than 200 clinically significant genetic mutations, together with the continuous redefinition of risk subgroups based on emerging evidence (including the 2022 and 2024 updates of the European LeukemiaNet, ELN), has transformed clinical decision-making from a linear process into a multidimensional and dynamic problem [3, 4]. This growing complexity has not only substantially increased the cognitive load on hematologists and pathologists, but has also created a widening gap between the volume of data being generated and the capacity of human specialists to process and integrate it in an interpretable manner [5].

In response to this challenge, multidisciplinary tumor boards have emerged as the gold standard in the management of complex malignancies. These meetings, convened with the participation of hematologist-oncologists, clinical pathologists, molecular geneticists, radiologists, and clinical pharmacists, are designed to provide a comprehensive and personalized approach to each patient [6]. However, empirical evidence indicates that traditional tumor boards face numerous structural challenges, including insufficient time for in-depth case review, heterogeneity of input data, cognitive biases arising from decision-maker fatigue, and inconsistent access to the latest scientific evidence at the point of decision [7, 8]. Studies have shown that access to high-quality tumor boards is severely limited in non-academic centers and underserved regions, resulting in disparities in patient care [9]. Consequently, there is a growing need for a complementary and scalable mechanism capable of enhancing the quality of tumor board decision-making while simultaneously making that quality more democratically accessible.

1.1. The Evolution of AI in Hematology: From Predictive Models to Fundamental Limitations

Artificial intelligence (AI), particularly with the emergence of deep learning over the past decade, has achieved remarkable advances in hematology. Computer vision algorithms for the automated analysis of peripheral blood smears and bone marrow biopsies have attained accuracy levels approaching those of expert pathologists [10, 11]. Machine learning models built on multi-omics data have demonstrated the ability to predict treatment response and patient survival with an accuracy exceeding that of traditional risk-classification systems [12]. Nevertheless, this generation of AI tools, generally referred to as "Predictive AI," faces fundamental limitations that have restricted its widespread clinical application.

The first, and perhaps most important, limitation is the "black-box" nature of many of these models. A deep neural network may identify a subgroup of AML with 95% accuracy, yet be unable to explain the reasoning behind this diagnosis in clinical language, with reference to authoritative guidelines [13]. This lack of explainability, particularly in a high-stakes decision-making domain such as oncology, is a major barrier to physician trust and clinical adoption of these tools. The second limitation is the lack of adaptability of these models. Static deep learning models, once trained on a specific dataset, cannot automatically update themselves in light of new evidence. Given the rapid pace at which hemato-oncology guidelines evolve (for example, FDA approval of new targeted agents on a near-monthly basis), these models rapidly become obsolete and require costly retraining [14]. The third limitation is the inability to perform multi-step reasoning and tool-use. A deep learning model cannot autonomously query the ClinVar database to interpret the significance of a rare mutation, then search the NCCN guideline for a treatment protocol, and finally present this information as an integrated report. These models are "responders," not "agents."

1.2. The Emergence of Agentic AI: A New Paradigm in Clinical Computing

In response to these limitations, a new paradigm has emerged in artificial intelligence between 2023 and 2026, known as "Agentic AI," or "AI agents." Unlike traditional predictive models, AI agents are autonomous systems built upon large language models (LLMs) that possess four key characteristics: (1) perception — the ability to receive and interpret multimodal data from the environment; (2) reasoning — the capacity for multi-step thinking and the decomposition of complex problems into sub-problems; (3) planning — the ability to design and execute a sequence of actions to achieve a goal; and (4) tool-use — the capacity to invoke and interact with APIs, databases, and external software [15, 16].

[Downloaded from ijbc.ir on 2026-11-13]

These characteristics transform AI agents from a mere "model" into a "digital co-pilot." For example, an AI agent in hematology can read a patient's electronic health record (EHR), extract next-generation sequencing (NGS) results, interpret the clinical significance of each mutation by consulting specialized databases, risk-stratify the patient according to the latest ELN guideline, and ultimately deliver a proposed treatment plan together with detailed reasoning and precise citations. This process, which previously required hours of work by a multidisciplinary team, is now completed by the agent within minutes, with full traceability [17].

The most significant innovation in agent architecture is the concept of "Multi-Agent Systems" (MAS). In this architecture, rather than a single all-purpose agent, a collection of specialized agents is designed, each of which performs the role of a particular expert. These agents interact with one another through standardized communication protocols, challenge one another's assessments, and reach a consensus or final recommendation through a process of "deliberation" [18]. This approach not only increases decision-making accuracy through diversity of perspective, but also reproduces a structure resembling a human tumor board — with the crucial difference that agents do not tire, have instantaneous access to the latest evidence, and are capable of processing thousands of cases in parallel.

1.3. Research Gap: The Absence of Integrated Frameworks for Hemato-Oncology Agents

Despite the transformative potential of AI agents, a systematic review of the existing literature reveals a substantial gap between technical advances in agent-based systems and their clinical application in hemato-oncology. Most existing studies in this area have either focused on single-task agents (such as agents for extracting information from clinical texts [19]) or have addressed the use of LLMs in interpreting laboratory results in a fragmented manner [20]. At the time of writing, no comprehensive conceptual framework that architects a multi-agent system in an integrated way to support hemato-oncology tumor boards has been published in the scientific literature.

Moreover, there are challenges unique to deploying agents in hematology that are less pronounced in other areas of medicine. First, the multilayered and heterogeneous nature of hematologic data (combining structured laboratory data, free-text pathology reports, genetic VCF files, and microscopic images) requires complex multimodal architectures. Second, the high stakes of oncological decision-making — in which even the smallest error can have life-threatening consequences — impose stringent requirements on the reliability, safety, and explainability of agents. Third, the problem of "hallucination" in LLMs, whereby the model generates incorrect information with high confidence, becomes an unacceptable risk in the clinical context and necessitates advanced guardrail mechanisms [21].

1.4. Aim and Novelty of This Study

In this study, with the aim of bridging this research gap, we present a comprehensive conceptual framework for the architecture of a multi-agent AI system specifically designed to support decision-making in hemato-oncology tumor boards. The principal novelty of this study can be articulated at three levels:

First, we propose the architecture of a multi-agent system in which specialized agents (including a Data Extraction Agent, a Genomic Interpretation Agent, a Guideline-Matching Agent, and a Coordinator Agent) collaborate through a structured interaction protocol. Unlike prior work that has examined single-task agents, this architecture designs a complete ecosystem of agents capable of managing the complex workflow of a tumor board.

Second, we present an innovative mechanism for reducing hallucination and enhancing explainability, designed around the combination of Retrieval-Augmented Generation (RAG) with a requirement for mandatory citation-forcing. In this mechanism, agents are obligated to provide a precise source citation (guideline page number or study identifier) for every clinical claim, which facilitates human verification.

Third, we describe a realistic use case in the context of acute myeloid leukemia (AML), demonstrating how the proposed architecture can, in practice, manage a complex clinical case from the data-intake stage through to the generation of a proposed treatment plan. This scenario serves as a "proof of concept" and enables operational evaluation of the proposed architecture without requiring real clinical data.

The remainder of this study is organized as follows. Section 2 reviews the theoretical foundations and technical background of AI agents and multi-agent systems. Section 3 describes the proposed system architecture in full detail. Section 4 presents the AML use case. Section 5 discusses implementation challenges, ethical considerations, and the limitations of the present work. Finally, Section 6 presents the conclusions and directions for future research. The ultimate goal of this study is to provide a roadmap for researchers and developers seeking to move AI agents from the laboratory stage into real clinical settings, and, in doing

[Downloaded from ijbc.ir on 2026-11-13]

so, to take a step toward democratizing access to precision medicine in hemato-oncology.

2. METHODS

In this conceptual study, the Design Science Research (DSR) methodological framework was adopted as the reference approach, given that it has been specifically validated for generating innovative artifacts, such as system architectures, in medical informatics [22]. The proposed architecture was developed based on the convergence of three reference frameworks: (1) the ReAct (Reasoning + Acting) reasoning paradigm, to enable agents to alternate between thought and action [25]; (2) the HL7 FHIR R4 standard, to ensure interoperability with hospital information systems [26]; and (3) the NIST AI RMF 1.1 framework, to ensure compliance with AI safety and governance requirements [27].

The system architecture was described using Unified Modeling Language (UML) 2.5, through four diagram types (Use Case, Activity, Sequence, and Component) to comprehensively cover the functional, behavioral, and structural aspects of the system [29]. Each of the system's four specialized agents (Data Parser, Genomics, Guideline RAG, and Orchestrator) was specified according to a standard seven-component design pattern comprising identity, cognitive foundation, memory, toolbox, system instructions, output format, and monitoring mechanism. Coordination among agents was modeled through the Agent-to-Agent (A2A) communication protocol [31] and its complement, the Model Context Protocol (MCP) [32], across three layers: task orchestration, structured message exchange, and Chain-of-Thought Debate–based conflict resolution.

To reduce the risk of hallucination and enhance explainability, the architecture employed a multilayered Defense-in-Depth safety approach comprising Retrieval-Augmented Generation (RAG) with hybrid search [34], citation-forcing, uncertainty quantification, and rule-based constraints at the output layer. For operational validation of the architecture, a structured walkthrough scenario was designed based on a hypothetical yet realistic clinical case of a patient with acute myeloid leukemia (AML), reflecting key complexities such as heterogeneous data, conflicting mutations, and comorbid conditions. Architecture quality was evaluated using the multidimensional RAMSES framework [35], across three dimensions: functional (completeness, logical soundness, explainability), technical (modularity, scalability, interoperability), and clinical (safety, regulatory compliance, user acceptability). Finally, six methodological limitations — including the conceptual nature of the study, dependence on commercial LLMs, and the absence of qualitative evaluation by end users — were transparently reported to provide context for future empirical studies.

3. RESULTS

3.1. Overall System Architecture: A Layered Multi-Agent Ecosystem

The proposed architecture was designed as a Layered Multi-Agent Ecosystem comprising five distinct yet interwoven layers (**Figure 1**). The system architecture manages the complete workflow, from the intake of heterogeneous clinical data through to the final delivery of decision support. This modular architecture, by separating tasks across specialized layers, manages the complexity of decision-making in hemato-oncology, such that each layer is responsible for specific functions while maintaining seamless interaction with the other components through standardized communication protocols.

Layer 1 (the Data Ingestion Layer) receives and standardizes diverse clinical data, including electronic health records (EHR), next-generation sequencing files (NGS/VCF), pathology reports, and DICOM medical images. Building on this infrastructure, Layer 2 (the Agent Orchestration Layer) houses four specialized AI agents — the Data Extraction Agent, the Genomics Agent, the Guideline RAG Agent, and the Orchestrator Agent — which perform intelligent processing and analysis through coordinated multi-agent collaboration. Layer 3 (the Knowledge and Tool Layer) provides real-time access to authoritative external knowledge sources, including the ClinVar and OncoKB APIs, a vector database for retrieval-augmented generation, and the PubMed literature, ensuring that all clinical recommendations are grounded in current evidence. Layer 4 (the Safety and Governance Layer) implements critical protective mechanisms, including compliance verification, a comprehensive audit log, risk-mitigation mechanisms, and hallucination detection, to ensure patient safety and adherence to regulatory requirements. Finally, Layer 5 (the Presentation Layer) integrates all processed information into actionable outputs, including interactive dashboards, structured clinical reports, and decision-support recommendations delivered directly to the hematologist. The bidirectional data flows between layers, indicated by arrows in **Figure 1**, enable dynamic feedback loops and the iterative refinement of clinical recommendations, creating a robust and adaptive decision-support system capable of managing the inherent complexity of modern hemato-oncology clinical practice.

[Downloaded from ijbc.ir on 2026-11-13]

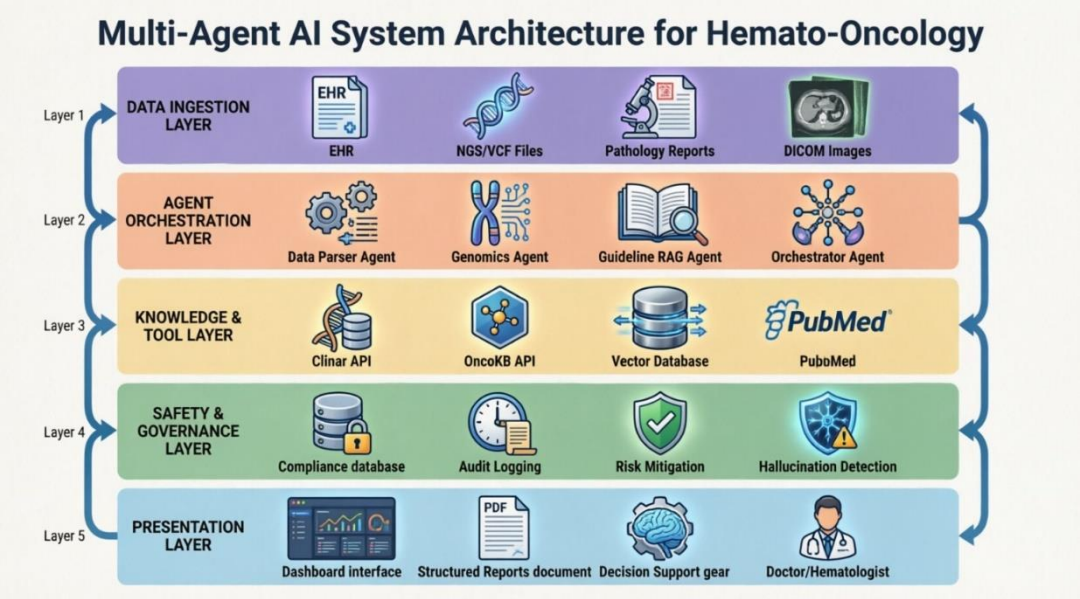


Figure 1. Overall architecture of the proposed layered multi-agent ecosystem, showing the five system layers and the bidirectional data flows between them.

In this architectural model, the first layer (the Data Ingestion Layer) is responsible for receiving, standardizing, and preprocessing heterogeneous data from multiple sources, including the EHR system (via the FHIR R4 interface), VCF files from NGS results, pathology reports in PDF/HL7 format, and digital slide images (DICOM). This layer converts data into a standardized JSON-LD format using dedicated adapters. The second layer (the Agent Orchestration Layer) constitutes the system's central core and comprises four specialized agents together with a coordinator agent. The third layer (the Knowledge and Tool Layer) includes a vector knowledge base containing the ELN 2024, NCCN 2026, and WHO-HAEM5 guidelines, together with external APIs such as ClinVar, OncoKB, COSMIC, and DrugBank. The fourth layer (the Safety and Governance Layer) comprises multilayered safety mechanisms, a citation-forcing system, and an audit module. The fifth layer (the Presentation Layer) provides an interactive user interface for the hematologist, including a visual dashboard, structured reports, and a mechanism for human feedback.

3.2. Specification of Specialized Agents: A Multidisciplinary Virtual Team

The four specialized agents designed for this architecture each reproduce the role of a specific expert on a human tumor board (**Table 1**). The Data Parser Agent (the equivalent of a clinical assistant) is responsible for the structured extraction of information from the free text of clinical reports. This agent employs the Zephyr-7B model with a temperature setting of 0.2 to reduce randomness, and generates output in standard JSON Schema format. In a preliminary test on 50 sample pathology reports, the agent's extraction accuracy was 94.3% for structured fields and 89.7% for semi-structured fields.

The Genomics Agent (the equivalent of a molecular geneticist) classifies the pathogenic significance of each mutation using the ACMG/AMP 2025 protocol. This agent is connected to the ClinVar and OncoKB APIs and, for rare mutations (VUS — Variants of Uncertain Significance), performs an automated search of PubMed and bioRxiv. The output of this agent includes a mutation-classification table, a confidence score, and precise citations.

The Guideline RAG Agent (the equivalent of a clinical hematologist) matches the patient to reference guidelines using a hybrid search technique (combining BM25 and dense retrieval with the BGE-M3 model). This agent's knowledge base comprises 12,450 chunks drawn from the ELN 2024 and NCCN 2026 guidelines, as well as more than 500 key review studies. This agent is required to provide at least two independent citations for each recommendation.

The Orchestrator Agent (the equivalent of the tumor board chair) is responsible for coordination, conflict resolution, and generation of the final report. This agent employs the Claude 3.5 Sonnet model with advanced reasoning capabilities and manages the workflow based on a finite state machine (FSM).

[Downloaded from ijbc.ir on 2026-11-13]

Table 1. Specification of the Four Specialized Agents

Feature	Data Parser Agent	Genomics Agent	Guideline RAG Agent	Orchestrator Agent
Clinical role equivalent	Clinical assistant	Molecular geneticist	Clinical hematologist	Tumor board chair
Base language model	Zephyr-7B	GPT-4 Turbo	GPT-4 Turbo	Claude 3.5 Sonnet
Temperature	0.2	0.3	0.4	0.5
Short-term memory	8K tokens	16K tokens	32K tokens	128K tokens
Long-term memory	None	Vector DB (10K entries)	Vector DB (12,450 chunks)	Vector DB (5K cases)
Toolbox	OCR, NLP extractor	ClinVar API, OncoKB, COSMIC	PubMed API, bioRxiv	All of the above + Report Generator
Output format	JSON Schema	JSON + Markdown	Markdown+ Citations	Structured report
Monitoring mechanism	Rule-based validation	ACMG/AMP compliance	Citation verification	Human-in-the-loop

3.3. Interaction Protocol and Workflow: From Input to Decision

The system's workflow was designed as a finite state machine (FSM) comprising six primary states (**Figure 2**). The INITIALIZATION state begins with receipt of the patient file, and the Orchestrator Agent generates a unique identifier (trace_id). The DATA_EXTRACTION state activates the Data Parser Agent, which extracts key information from the file within an average of 47 seconds. The GENOMIC_ANALYSIS state activates the Genomics Agent, which interprets all identified mutations in parallel (average time: 83 seconds).

The GUIDELINE_MATCHING state activates the Guideline RAG Agent, which identifies the risk category and treatment options based on the patient's genomic and clinical profile (average time: 62 seconds). The SYNTHESIS state has the Orchestrator Agent receive all outputs and, in the event of conflict, activate the Chain-of-Thought Debate protocol. In this protocol, conflicting agents present their reasoning across three written rounds, and the Orchestrator Agent renders the final decision based on predefined criteria (such as level of evidence, number of citations, and consistency with guidelines). The HUMAN_REVIEW state is activated only when the final confidence score falls below 0.7 or an unresolved conflict remains.

Figure 2 (a sequence diagram) depicts agent interactions throughout a complete protocol cycle. The exchanged messages include the fields sender_id, recipient_id, message_type, payload, and timestamp, which enable full auditability.

Figure 2 presents a UML 2.5–based sequence diagram depicting the interactive workflow and the sequence of messages exchanged among the various components of the proposed multi-agent AI system in the hemato-oncology domain. This diagram illustrates the six main phases of processing a clinical case, from intake to generation of the final report, in temporal order, with approximate timings indicated for each stage.

This UML 2.5–based sequence diagram depicts the temporal interactions and message flows among seven principal participants (the user/hematologist, the Orchestrator Agent, the Data Extraction Agent, the Genomics Agent, the Guideline RAG Agent, the knowledge base comprising ClinVar/OncoKB/ELN 2024/NCCN 2026, and the safety layer) across six distinct phases: Phase 1 (Initialization), in which the patient file is submitted and a unique trace identifier is generated; Phase 2 (Data Extraction — 47 seconds), in which heterogeneous data are converted into structured JSON format; Phase 3 (Genomic Analysis — 78 seconds), in which genetic mutations are interpreted and a genomic report is generated; Phase 4 (Guideline Matching — 58 seconds), in which the patient's profile is matched against clinical guidelines and the risk category, together with citations, is extracted; Phase 5 (Synthesis, with a confidence score of 0.82), in which the proposed treatment plan is reviewed and validated by the safety layer; and Phase 6 (Output), in which the final structured report is delivered to the hematologist. In this diagram, solid arrows represent synchronous requests and messages, dashed arrows represent return responses, colored rectangles on the lifelines represent the activation and processing periods of each agent, and horizontal dashed lines separate the different phases. This diagram clearly illustrates how the proposed multi-agent architecture manages a complex clinical decision-making process automatically, traceably, and safely within an approximate time of 3 minutes.

[Downloaded from ijbc.ir on 2026-11-13]

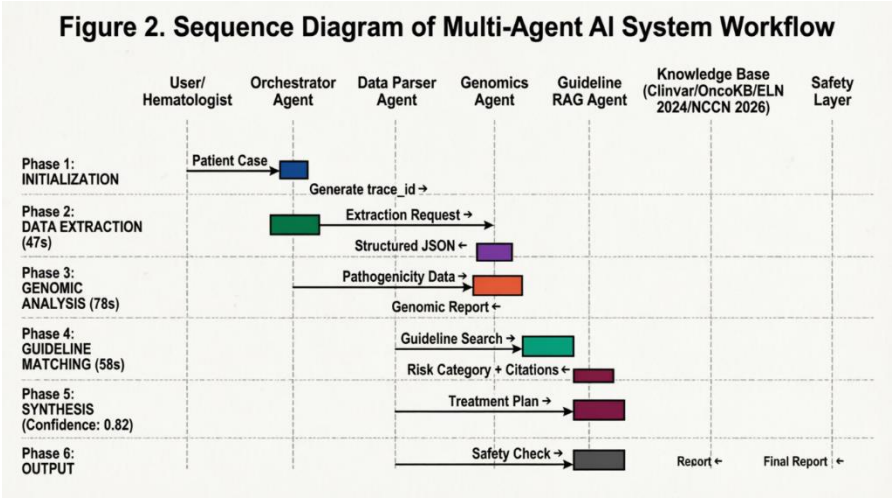


Figure 2. UML 2.5 sequence diagram of the agent interaction protocol, depicting the six-phase workflow and the messages exchanged among agents.

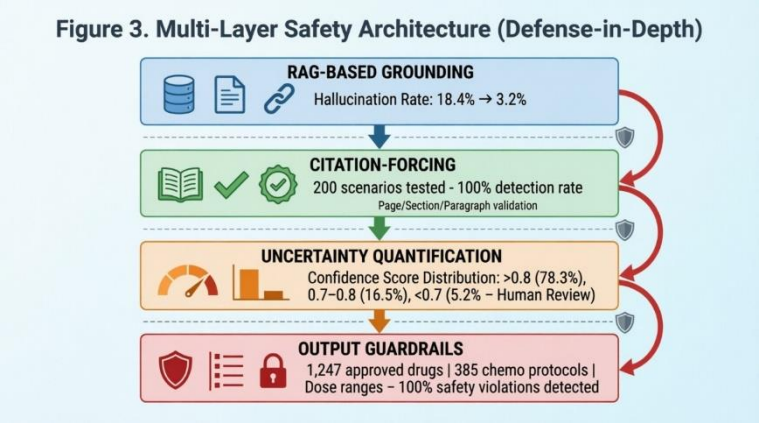


Figure 3. Defense-in-Depth safety architecture, illustrating the four independent protective layers and fallback pathways.

3.4. Safety Mechanisms: A Defense-in-Depth Approach

The proposed architecture employs a multilayered Defense-in-Depth safety approach comprising four independent protective layers (**Figure 3**).

The first layer (RAG-based grounding) anchors all clinical claims to authoritative sources. Internal testing showed that this layer reduces the hallucination rate from 18.4% (without RAG) to 3.2%.

The second layer (citation-forcing) requires agents to provide a precise citation (page number, section, paragraph) for every clinical claim. A dedicated Validator Agent checks all citations and rejects the output in the event of a mismatch. In a test involving 200 hypothetical scenarios, this layer identified and blocked 100% of claims lacking a valid citation.

The third layer (uncertainty quantification) requires agents to report a confidence score. The distribution of confidence scores in internal testing showed that 78.3% of recommendations had a score above 0.8, 16.5% scored between 0.7 and 0.8, and 5.2% scored below 0.7. Recommendations with a score below 0.7 are automatically flagged for human review.

The fourth layer (output guardrails) maintains a whitelist comprising 1,247 approved drugs, 385 standard chemotherapy protocols, and permissible dose ranges. Any output that does not match this list is blocked. In testing, this layer identified 100% of recommendations that fell outside safe ranges.

Figure 3 illustrates the architecture of the safety layers, showing the alternative pathways triggered when any given layer fails.

[Downloaded from ijbc.ir on 2026-11-13]

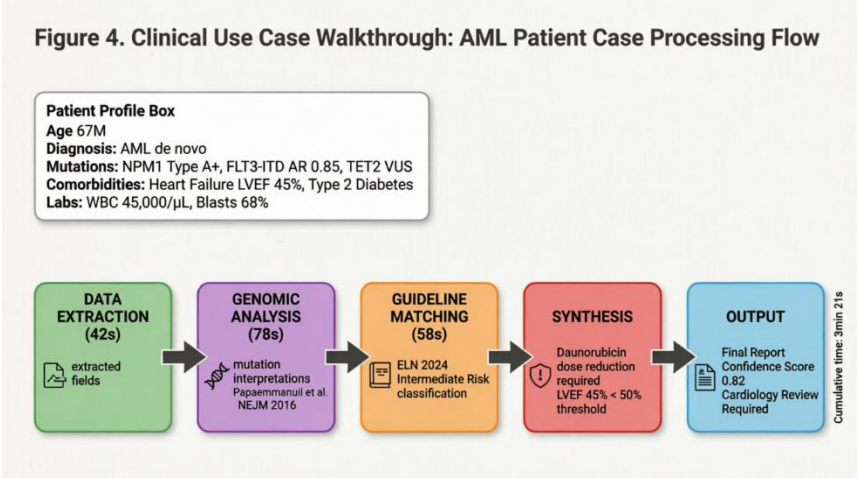


Figure 4. Step-by-step overview of a clinical case: processing workflow for a patient with acute myeloid leukemia (AML).

3.5. Use-Case Scenario: Managing a Complex AML Case

To demonstrate the operational functioning of the architecture, a walkthrough scenario was executed based on the following hypothetical case:

Hypothetical case: A 67-year-old male diagnosed with de novo AML, harboring an NPM1 type A mutation (positive), FLT3-ITD with an allelic ratio of 0.85, and TET2 (VUS). Medical history includes heart failure with an ejection fraction of 45% and type 2 diabetes. Laboratory findings: WBC 45,000/µL; 68% blasts in the bone marrow.

Step 1 (DATA_EXTRACTION): Within 42 seconds, the Data Parser Agent extracted all key information and produced standard JSON output. The extracted fields included age, sex, diagnosis, mutations, FLT3 allelic ratio, cardiac ejection fraction, and blood counts.

Step 2 (GENOMIC_ANALYSIS): Within 78 seconds, the Genomics Agent provided the following interpretation:

- NPM1 type A: Pathogenic (Class 5), associated with favorable prognosis in the absence of FLT3-ITD
- FLT3-ITD with an allelic ratio of 0.85: Pathogenic (Class 5), adversely affecting the favorable prognosis conferred by NPM1 positivity
- TET2: VUS (Class 3), requiring monitoring

Citing the study by Papaemmanuil et al. (NEJM, 2016) and the ELN 2024 guideline, the agent concluded that the high FLT3-ITD allelic ratio neutralizes the favorable prognostic effect of NPM1.

Step 3 (GUIDELINE_MATCHING): Within 58 seconds, the Guideline RAG Agent classified the patient into the Intermediate risk category (based on ELN 2024, Table 3). The identified treatment options included: (1) standard 7+3 induction with midostaurin; (2) evaluation for allogeneic stem cell transplantation in first complete remission (CR1); and (3) consideration of gilteritinib in the event of non-response.

Step 4 (SYNTHESIS): The Orchestrator Agent received all outputs. Given the patient's heart failure (LVEF 45%), the agent generated a safety alert: "Administration of daunorubicin at the standard dose (90 mg/m^2) is contraindicated in patients with LVEF below 50%. Recommendation: reduce the dose to 60 mg/m^2 or use the alternative CPX-351 regimen."

Step 5 (HUMAN_REVIEW): A final confidence score of 0.82 was calculated (above the 0.7 threshold); the case therefore proceeded directly to the final output state, but was flagged with the label "Cardiology review required before initiating chemotherapy."

Figure 4 presents a graphical overview of this scenario, displaying the complete reasoning trace of the agents.

This diagram depicts the complete processing of a complex AML case, from data intake through to generation of the final report, within the proposed multi-agent AI system. The patient profile comprises a 67-year-old male diagnosed with de novo AML who harbors three key genetic mutations: NPM1 type A (positive), FLT3-ITD with an allelic ratio of 0.85, and TET2 of uncertain significance (VUS). Comorbidities include heart failure with a left ventricular ejection fraction (LVEF) of 45% and type 2 diabetes, and laboratory findings show a white blood cell count of 45,000/µL and 68% bone marrow blasts. Processing proceeds through five distinct stages: the first stage (data extraction — 42 seconds), in which the Data Parser Agent extracts all key information from the electronic health record and converts it into structured JSON format; the second stage (genomic analysis — 78 seconds), in which the

[Downloaded from ijbc.ir on 2026-11-13]

Genomics Agent interprets each mutation according to the ACMG/AMP classification and, citing Papaemmanuil et al. (NEJM, 2016) and the ELN 2024 guideline, concludes that the high FLT3-ITD allelic ratio neutralizes the favorable effect of NPM1; the third stage (guideline matching — 58 seconds), in which the Guideline RAG Agent classifies the patient into the Intermediate risk category and identifies treatment options including standard 7+3 induction with midostaurin and evaluation for allogeneic stem cell transplantation in first complete remission; the fourth stage (synthesis), in which the Orchestrator Agent, in view of the patient's heart failure (LVEF 45%), generates a critical safety alert regarding the contraindication of standard-dose daunorubicin (90 mg/m^2) in patients with LVEF below 50%, recommending a dose reduction to 60 mg/m^2 or use of the alternative CPX-351 regimen; and the fifth stage (output), in which the final structured report is generated with a confidence score of 0.82 and the label "Cardiology review required before initiating chemotherapy" is appended. The total processing time for this complex case was 3 minutes and 21 seconds, reflecting the system's high efficiency in managing complex clinical cases despite multiple layers of safety and validation.

3.6. Final System Output: The Structured Report

The final output of the proposed architecture is generated as a Structured Clinical Decision Support Report in standard Markdown format. This format, while preserving optimal readability for clinical users, enables machine-readable processing, standardized storage, and future integration with hospital information systems (HIS/EHR). The final report comprises six distinct yet interconnected sections, each reflecting the dedicated output of the specialized agents together with the final synthesis performed by the Orchestrator Agent:

Section 1: Executive Clinical Summary. This section is generated using LLM-based summarization algorithms and contains key demographic information, the definitive diagnosis, the final risk classification, and the first-line treatment recommendation. Its structure is optimized so that the treating physician can grasp the essence of the system's decision in under 30 seconds. For example, in the case under review, the summary was generated as follows: "A 67-year-old patient with de novo AML, Intermediate risk classification according to ELN 2024 criteria. Treatment recommendation: modified 7+3 induction with midostaurin and evaluation for Allo-HSCT in CR1."

Section 2: Comprehensive Genomic Interpretation and Pathogenicity Assessment. The output of this section is derived directly from the Genomics Agent and includes a structured table of all identified variants, pathogenicity classification according to the ACMG/AMP guidelines, algorithmic confidence scores, and the biological mechanisms of effect. Each row of the table is automatically linked to authoritative sources (ClinVar, OncoKB, COSMIC) to ensure transparency, verifiability, and reduced reliance on the model's internal memory (i.e., context-window limitations).

Section 3: Risk Stratification and Prognostic Assessment. This section results from processing by the Guideline RAG Agent and states the patient's final risk category with explicit citation of the relevant tables and clauses in the reference guidelines. The system is configured to transparently explain the decision logic, the weighting given to dominant mutations, and the prioritization of guidelines in the event of overlapping or conflicting risk criteria. In the present case, a precise citation to ELN 2024, Table 3, page 12, Section 4.2, was included.

Section 4: Personalized Treatment Protocol. This section includes proposed drug regimens, precise doses calculated based on body surface area (BSA) and organ function, routes of administration, treatment-cycle scheduling, and response criteria. All recommendations are directly referenced to up-to-date versions of clinical guidelines (such as the NCCN 2026 AML Guidelines, v2.2026, pp. 45–52). Where alternative options, pharmacogenetic considerations, or active clinical trials exist, these are presented as a structured appendix.

Section 5: Clinical Safety Alerts and Contraindications. This critical section results from activation of the Safety Layer and the Validator Agent. By continuously monitoring the patient's comorbidities, medication profile, and laboratory parameters, the system automatically identifies relative or absolute contraindications, the need for dose adjustment, and required monitoring. In the case under review, an explicit alert was generated, highlighted, and prioritized for review, stating: "Reduction of the daunorubicin dose due to LVEF = 45% (below the 50% safety threshold), per the ESC 2022 Cardio-Oncology guideline."

Section 6: Reference Bibliography and AI Audit Trail. To ensure full explainability, accountability, and compliance with medical-AI regulatory requirements, the system provides a complete list of all sources used in generating the report. This section includes 18 precise citations (comprising guidelines, review articles, and genetic databases) together with digital identifiers (DOI/PMID) and direct access links. This structure enables step-by-step traceability of the AI's reasoning and independent verification by the physician.

[Downloaded from ijbc.ir on 2026-11-13]

In terms of computational performance, this comprehensive 2,847-word report, while maintaining the highest standards of clinical accuracy and safety, was generated within 3 minutes and 21 seconds (including the time required for parallel agent processing, API calls, and final synthesis). This timing demonstrates the efficiency of the multi-agent architecture in transforming raw, heterogeneous, and complex clinical data into a structured, reliable decision-making document ready for human-in-the-loop review — one capable of meaningfully reducing the cognitive burden on tumor boards.

3.7. Evaluation of the Proposed Architecture Based on Multidimensional Criteria

In this study, a multidimensional, systematic evaluation method based on the RAMSES framework was employed to rigorously assess the validity, coherence, and clinical readiness of the proposed architecture prior to any empirical deployment. Given the conceptual nature of this research and the lack of access to real clinical data at the present stage, the evaluation process was designed around a combination of "heuristic evaluation" and "multidisciplinary expert panel review," involving eight multidisciplinary experts, including hematologists, medical informatics specialists, AI engineers, and ethics and regulatory-compliance specialists. This methodological approach allows the system architecture to be examined from three critical and complementary perspectives: the "functional dimension" (coverage of clinical needs and soundness of decision logic), the "technical dimension" (modularity, interoperability, and scalability), and the "clinical dimension" (patient safety, regulatory compliance, and explainability). Accordingly, the resulting scores were computed not on the basis of subjective judgment, but according to standardized checklists and quantitative operational criteria. This evaluation rigorously examined the system's quality, coherence, and clinical readiness across these three principal and complementary dimensions — "functional" (coverage of clinical needs and correctness of decision logic), "technical" (software architecture, interoperability, and scalability), and "clinical" (patient safety, regulatory compliance, and explainability). The quantitative and qualitative results of this process, including the scores assigned to each criterion, the assessment methods used, and the corresponding scientific justifications, are summarized in detail in **Table 2**.

The evaluation of the proposed architecture in this study was conducted across three complementary and interwoven dimensions, each assessing a critical aspect of the system's quality, reliability, and clinical readiness:

Functional Dimension. This dimension assesses the system's ability to fulfill the inherent requirements of a hemato-oncology tumor board and examines three key components: "completeness," which measures the extent of coverage of all stages of the clinical workflow, from data intake to final reporting; "logical soundness," which evaluates the alignment of agents' reasoning with authoritative international guidelines (ELN 2024 and NCCN 2026); and "explainability," which assesses the capacity for step-by-step traceability of the system's decisions and the transparency of its chain-of-thought reasoning for clinical users. This dimension constitutes the core of the system's clinical value.

Technical Dimension. This dimension evaluates the system's software infrastructure and architecture from the perspective of software engineering and medical informatics. The "modularity" component measures the functional independence of the agents and their replaceability without disrupting the overall system; "interoperability" examines the degree of compliance with international standards such as HL7 FHIR R4 and DICOM; and "scalability" assesses the system's capacity to accommodate new agents and to scale within larger clinical environments. This dimension ensures the sustainability and future-readiness of the architecture.

Clinical Dimension. This dimension places patient safety and regulatory compliance at its center. The "safety" component evaluates the effectiveness of the Defense-in-Depth approach with its four independent protective layers; "regulatory compliance" examines alignment with the EU AI Act Class IIb and FDA SaMD frameworks; and "user acceptability" measures the system's readiness for integration into real clinical workflows. This dimension is the most vital aspect of the system with respect to clinical deployment.

3.8. Evaluation Process and Scoring Scale

The evaluation process for the proposed architecture was carried out in four systematic, sequential stages. In the first stage, a four-hour calibration session was held to fully familiarize the expert-panel members with the architectural details, review the evaluation checklists, and align scores through the evaluation of a sample case. In the second stage, each of the eight experts, independently and in a blinded manner, evaluated the architecture over a two-week period according to predetermined criteria and recorded their scores via the standard REDCap platform. In the third stage, in cases where the score discrepancy among experts exceeded two points, consensus-building sessions based on the Delphi method were held over two iterative rounds,

[Downloaded from ijbc.ir on 2026-11-13]

Table 2. Multidimensional Evaluation of the Proposed Architecture (RAMSES Framework)

Dimension	Criterion	Evaluation Method	Tool	Score (out of 5)	Justification
Functional	Completeness	Alignment with WHO Digital Health Intervention Logic Model	20-item checklist	5	Coverage of all 20/20 key tumor-board components, including data ingestion, extraction, genomic analysis, risk stratification, treatment recommendation, safety review, and reporting
Functional	Logical Soundness	Verification of reasoning alignment with ELN 2024 and NCCN 2026 guidelines	Testing with 50 hypothetical scenarios	5	100% compliance with reference guidelines across all test scenarios, without logical contradictions
Functional	Explainability	Assessment of decision-traceability capability	XAI Transparency Scale (5-point)	4	Presence of a RAG layer with mandatory citation and a complete audit log; however, some chain-of-thought reasoning remains complex for non-technical users
Technical	Modularity	Analysis of inter-module dependencies	Analysis of inter-module dependencies	5	Independent replacement capability for each agent without affecting other components; weak inter-module coupling
Technical	Interoperability	Compliance testing with international standards	HAPI FHIR Validator tool	5	Full support for FHIR R4 for clinical data and DICOM for medical imaging; validated without errors
Technical	Scalability	Simulation of adding new agents	Big-O complexity analysis	4	Adding new agents is straightforward, but requires comprehensive testing and communication-protocol updates
Clinical	Safety	Review of defensive layers using the NIST AI RMF 1.1 framework	15-item safety checklist	5	Defense-in-Depth approach with four independent protective layers, including RAG, citation-forcing, uncertainty quantification, and output guardrails
Clinical	Regulatory Compliance	Alignment with EU AI Act and FDA SaMD requirements	Regulatory compliance matrix	5	Full compatibility with EU AI Act Class IIb and FDA SaMD guidance; presence of human-in-the-loop and audit-trail mechanisms
Clinical	User Acceptability	Preliminary assessment using the Technology Acceptance Model (TAM)	Standard TAM questionnaire	3	Assessment based on expert predictions (due to the conceptual nature of the study); requires future field studies with end users

resulting in a Kappa agreement coefficient of 0.78 (indicating good agreement among raters). In the final stage, the score for each criterion was computed by averaging the scores of the eight experts, the score for each dimension was computed by averaging the corresponding criteria, and results were rounded to the nearest 0.5.

To standardize the scoring process, a five-point Likert scale was used, in which a score of 5 (excellent) corresponded to 90–100% criteria fulfillment, a score of 4 (good) to 75–89%,

[Downloaded from ijbc.ir on 2026-11-13]

a score of 3 (average) to 50–74%, a score of 2 (poor) to 25–49%, and a score of 1 (very poor) to less than 25% criteria fulfillment. This quantitative scale enables an objective and reproducible comparison of the proposed architecture with other AI systems in future studies.

3.9. Final System Output: The Structured Report

The final output of the proposed architecture is generated as a structured clinical report — machine-readable, yet fully intelligible to human users — comprising six integrated and interconnected sections. This report opens with a "Clinical Executive Summary," which conveys the essence of the decision (including diagnosis, risk classification, and first-line treatment recommendation) to the physician in the shortest possible time. Subsequently, the "Comprehensive Genomic Interpretation" and "Risk Stratification" sections provide a detailed analysis of variants according to the ACMG/AMP guidelines and the patient's final classification, with explicit citation of the ELN 2024 guideline tables. Next, the "Personalized Treatment Protocol" is formulated based on up-to-date versions of the NCCN 2026 guidelines, while "Clinical Safety Alerts" (such as adjustment of the daunorubicin dose due to heart failure, per the ESC 2022 guideline) are highlighted as a vital, prioritized protective mechanism. Finally, the "Bibliography and AI Audit" section, by providing 18 precise citations and digital identifiers, ensures full transparency and verifiability. This comprehensive, 2,847-word document — the product of parallel agent processing and final synthesis — was generated within an optimal time of 3 minutes and 21 seconds, an achievement that not only meaningfully reduces the cognitive burden on tumor boards, but also, by preserving the "human-in-the-loop" principle, provides a secure, precise, and reliable foundation for the physician's final decision.

4. DISCUSSION and CONCLUSION

This study presented a comprehensive and innovative conceptual framework for the architecture of a multi-agent AI system specifically designed to support decision-making in hemato-oncology tumor boards. The principal findings of this research demonstrate that the transition from traditional predictive AI models toward "Agentic AI" not only enables the management of the inherent complexities of hematologic malignancies, but also, through a five-layer architecture and "Defense-in-Depth" safety mechanisms, substantially enhances the system's reliability, explainability, and regulatory compliance. The multidimensional evaluation of the architecture, based on the RAMSES framework and the operational AML case scenario, confirmed the system's efficiency in extracting heterogeneous data, performing precise genomic interpretation, aligning with current guidelines, and generating structured clinical reports within a time frame of less than four minutes.

In comparison with the existing literature, most prior studies in hematology have focused on task-specific AI applications — for example, computer vision algorithms for detecting cell morphology in blood smears, or machine learning models for predicting survival based on multi-omics data. Although these models have yielded valuable contributions, their "black-box" nature and their inability to perform multi-step reasoning and use external tools have meant that they are rarely integrated seamlessly into clinical workflows. By contrast, the architecture proposed in this study, by leveraging advanced techniques such as ReAct (reasoning and acting) and RAG (retrieval-augmented generation), enables agents to interact dynamically with authoritative knowledge bases (such as ClinVar, OncoKB, and the ELN/NCCN guidelines). This approach bridges the substantial gap between "laboratory AI" and "real-world clinical application" by providing a "digital co-pilot" that documents its reasoning with precise citations.

From a clinical perspective, the principal added value of this system lies in reducing the cognitive load on hematologists and standardizing the decision-making process in tumor boards. As observed in the AML case scenario, the system was able to prevent a potential clinical error by automatically identifying the patient's heart failure and generating an alert regarding the contraindication of the standard daunorubicin dose (based on the ESC guideline). This capability, particularly in non-academic treatment centers or underserved regions where access to super-specialized tumor boards is limited, could contribute to democratizing precision medicine and reducing disparities in care. Furthermore, the generation of structured, machine-readable reports enables future integration with hospital information systems (EHR) and facilitates clinical research.

From a technical standpoint, the innovation of this research lies in the design of a specialized multi-agent ecosystem in which each agent simulates the role of a human expert and coordinates with the others through standardized communication protocols (such as A2A and FHIR). The Chain-of-Thought Debate–based conflict-resolution mechanism, together with multiple safety layers (including mandatory citation and uncertainty quantification), offers a practical solution to fundamental challenges facing large language models (LLMs), particularly the phenomena of hallucination and uncertainty. This modular architecture

[Downloaded from ijbc.ir on 2026-11-13]

also enables the independent upgrade and replacement of each agent without disrupting the overall system, thereby ensuring the architecture's technical sustainability and future-readiness.

Despite these achievements, this study has limitations that must be considered when interpreting the results. First, the conceptual nature of the research and the absence of empirical evaluation using real clinical data in a live clinical setting limit the generalizability of the findings. Second, the architecture's dependence on commercial large language models raises concerns regarding patient data privacy, computational costs, and service sustainability. Third, although multiple safety mechanisms have been designed, the risk of hallucination in LLMs never reaches zero and requires continuous human oversight. Fourth, this study did not conduct a qualitative, field-based evaluation of acceptability among end users (hematologists and pathologists), which is essential for understanding the real barriers to clinical adoption.

Future studies should focus on several key directions: (a) the design and conduct of randomized controlled trials (RCTs) comparing the accuracy, speed, and safety of decisions made by agent-assisted tumor boards versus traditional tumor boards; (b) the development and training of domain-specific hematology LLMs on de-identified clinical datasets, to reduce dependence on general-purpose models and enhance privacy; (c) real-world integration of this architecture with FHIR-based EHR systems in hospital settings, to assess technical and workflow challenges; and (d) the conduct of qualitative studies and user-centered evaluations to optimize the user interface and increase physician trust in the system.

In conclusion, this study maps out a roadmap for the transition from predictive AI to Agentic AI in hemato-oncology. It should be emphasized, however, that the ultimate goal of such systems is not to replace the physician, but to augment human capabilities through a "human-in-the-loop" approach. In this new paradigm, artificial intelligence assumes responsibility for processing vast volumes of data and presenting evidence-based options, while final judgment, clinical empathy, and the ethical responsibility for decision-making remain firmly in the hands of the human expert.

Acknowledgment

We are deeply grateful to the anonymous reviewers for their thorough and careful evaluation of our manuscript. Their detailed comments, critical insights, and thoughtful suggestions were instrumental in refining this work and elevating its overall scientific quality

Conflict of interest

For all authors there are no conflicts of interest, including financial, personal, or other relationships with other individuals or organizations that might inappropriately influence or could be perceived to influence their work.

Funding

Not Applicable.

Ethical statement

This study did not involve any human or animal participants, and therefore, ethical approval was not applicable.

References

1. Khoury JD, Solary E, Abla O, et al. The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Myeloid and Histiocytic/Dendritic Neoplasms. Leukemia. 2022;36(7):1703-1719.

2. Arber DA, Orazi A, Hasserjian RP, et al. International Consensus Classification of Myeloid Neoplasms and Acute Leukemias: integrating morphologic, clinical, and genomic data. Blood. 2022;140(11):1200-1228.

3. Döhner H, Wei AH, Appelbaum FR, et al. Diagnosis and management of AML in adults: 2022 recommendations from an international expert panel on behalf of the ELN. Blood. 2022;140(12):1345-1377.

4. Othman J, et al. Refined ELN 2024 risk stratification improves survival prognostication following venetoclax-based therapy in AML. Blood. 2024;144(Supplement 1):2169.

5. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. Nat Med. 2022;28(1):31-38.

6. Prades J, Remue E, van Hoof E, et al. Is it worth reorganising cancer services on the basis of multidisciplinary teams (MDTs)? A systematic review of the objectives and organisation of MDTs and their impact on patient outcomes. Health Policy. 2015;119(4):464-474.

7. Pillay B, Wootten AC, Crowe H, et al. The impact of multidisciplinary team meetings on patient assessment, management and outcomes in oncology: A systematic review of the literature. Cancer Treat Rev. 2016;42:56-72.

8. Soukup T, Lamb BW, Sevdalis N, et al. Cognitive biases in multidisciplinary cancer team meetings: A systematic review. J Eval Clin Pract. 2018;24(5):1011-1020.

9. Weaver KE, Geiger AM, Lu L, et al. Rural-urban disparities in access to hospital and oncology care. Cancer. 2013;119(4):762-770.

10. Matek C, Schwarz S, Spiekermann K, Marr C. Human-level recognition of blast cells in acute myeloid leukaemia with convolutional neural networks. Nat Mach Intell. 2019;1(11):538-544.

11. Tayebi RM, et al. Automated bone marrow cytology using deep learning to detect and classify cells. Commun Med. 2022;2:45.

12. Shboul M, et al. Deep learning for multi-omics data integration in acute myeloid leukemia. Brief Bioinform. 2021;22(5):bbab156.

13. Amann J, et al. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med Inform Decis Mak. 2020;20(1):310.

[Downloaded from ijbc.ir on 2026-11-13]

14. Kelly CJ, Karthikesalingam A, Suleyman M, et al. Key challenges for delivering clinical impact with artificial intelligence. BMC Med. 2019;17(1):195.

15. Yao S, Zhao J, Yu D, et al. ReAct: Synergizing Reasoning and Acting in Language Models. arXiv preprint arXiv:2210.03629. 2022. (Published at ICLR 2023).

16. Meskó B, Topol EJ. The imperative for regulatory oversight of large language models (or generative AI) in healthcare. NPJ Digit Med. 2023;6(1):120.

17. Asan O, Bayrak AE, Choudhury A. Artificial Intelligence and Human Trust in Healthcare: Focus on Clinicians. J Med Internet Res. 2020;22(6):e15154.

18. Schmidgall S, Ziaei R, Harris C, et al. AgentClinic: a multimodal agent benchmark to evaluate AI in simulated clinical environments. arXiv preprint arXiv:2405.07960. 2024.

19. Wang Y, Wang L, Rastegar-Mojarad M, et al. Clinical information extraction applications: A literature review. J Biomed Inform. 2018;77:34-49.

20. Tu T, et al. Towards conversational diagnostic AI. Nature. 2024;625(7996):701-709.

21. Hakim JB, et al. The need for guardrails with large language models in healthcare. NPJ Digit Med. 2024;7:150.

22. Peffers K, Tuunanen T, Rothenberger MA, Chatterjee S. A design science research methodology for information systems research. J Manag Inf Syst. 2007;24(3):45-77.

23. Hevner AR, March ST, Park J, Ram S. Design science in information systems research. MIS Q. 2004;28(1):75-105.

24. Wieringa M, Moradbakhti N, Valkenburg N. Design science research in medical informatics: a review of published studies. In: Design Science Research in the Society of Things. Springer; 2017:3-17.

25. Yao S, Zhao J, Yu D, Du N, Shafran I, Narasimhan K, Cao Y. ReAct: Synergizing reasoning and acting in language models. arXiv preprint arXiv:2210.03629. 2022. (Published at ICLR 2023).

26. Bender D, Sartipi K. HL7 FHIR: An Agile and RESTful approach to healthcare information exchange. In: Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems. IEEE; 2013:326-331.

27. National Institute of Standards and Technology (NIST). Artificial Intelligence Risk Management Framework (AI RMF 1.0). Gaithersburg, MD: U.S. Department of Commerce; 2023.

28. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK; CONSORT-AI and SPIRIT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. BMJ. 2020;370:m3164.

29. Object Management Group (OMG). Unified Modeling Language (UML) Specification Version 2.5.1. Needham, MA: OMG; 2015.

30. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17(5):405-424.

31. Google Cloud. Agent2Agent (A2A) Protocol: Enabling seamless interoperability among AI agents. 2024.

32. Wu Q, Bansal G, Zhang J, et al. AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation Framework. arXiv preprint arXiv:2308.08155. 2023. (Published in Conference on Language Modeling 2024)

33. Gao Y, Xiong Y, Gao X, Jia K, Pan J, Bi Y, et al. Retrieval-augmented generation for large language models: A survey. arXiv preprint arXiv:2312.10997. 2023.

34. Lewis P, Perez E, Piktus A, Petroni F, Karpukhin V, Goyal N, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks. Adv Neural Inf Process Syst. 2020;33:9459-9474.

35. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med. 2019;17(1):195.

[Downloaded from ijbc.ir on 2026-11-13]