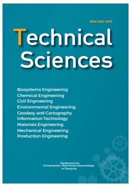
Technical Sciences, 2025, 28, 309–342

UNIWERSYTET
WARMIŃSKO-MAZURSKI
W OLSZTYNIE

Yearbook peer-reviewed scientific journal
ISSN 1505-4675
e-ISSN 2083-4527
TECHNICAL SCIENCES
Homepage: https://czasopisma.uwm.edu.pl/index.php/ts/



DOI: https://doi.org/10.31648/ts.12176

A PROPOSED REFERENCE MODEL FOR THE DEPLOYMENT OF AN INTEGRATED AI SYSTEM IN A LARGE ONCOLOGY CENTER UNDER THE EU AI ACT AND MDR PART I: STRATEGIC & OPERATIONAL FRAMEWORK

Mateusz Dąbkowski[1], ***Dominik Wawrzuta***[2], ***Ewelina Żarłok***[3],
Andrzej Jankowski[4], ***Lech Polkowski***[5], ***Andrzej Skowron***[6],
Piotr Artiemjew[7]

[1]ORCID: 0000-0002-5862-9747
[2]ORCID: 0000-0002-0846-8906
[3]ORCID: 0000-0002-9949-3460
[4]ORCID: 0000-0002-0725-6354
[5]ORCID: 0000-0002-6990-3959
[6]ORCID: 0000-0002-5271-6559
[7]ORCID: 0000-0001-5508-9856
[1,2]Maria Skłodowska-Curie National Research Institute of Oncology (NIO-PIB), Poland
[3]Revelva Concept, Poland
[4,5,7]University of Warmia and Mazury in Olsztyn (UWM), Poland
[6]Polish Academy of Sciences (PAN), Poland

Received 16 December 2025, accepted 22 December 2025, available online 31 December 2025

Keywords: oncology, integrated AI platform, reference architecture, reference deployment pathway, AI governance, auditability, EU AI Act, MDR, SaMD, human-in-the-loop (HITL), retrieval-augmented generation (RAG), GraphRAG.

Abstract

Large-scale deployment of AI in oncology is constrained less by standalone algorithmic performance than by system-level safety, accountability, interoperability, and regulation-aware governance. Grounded in approximately one year of practical pre-deployment work within the

Correspondence: Andrzej Jankowski, Katedra Metod Matematycznych Informatyki, Wydział Matematyki i Informatyki, Uniwersytet Warmińsko-Mazurski w Olsztynie, ul. Słoneczna 54, 10-710 Olsztyn, e-mail: andrzej.jankowski@uwm.edu.pl

OnkoBot project, this paper specifies a deployment- and governance-first reference model for integrated oncology AI platforms under the EU AI Act and the Medical Device Regulation (MDR).

The paper introduces Architecture for Medical AI Collaboration (AMAC), an implementation-neutral, system-level envelope that enforces strict online/offline separation between clinical operation and model/knowledge learning and evolution, gate-controlled releases via a Clinical Governance Gateway (CGG) with explicit human-in-the-loop (HITL) escalation, and tamper-evident auditability across clinical, technical, and interoperability boundaries. AMAC is anchored by the Community of Collaborative Evolving Medical Assistants (CEMA), a supervised multi-agent computational core that performs coordinated clinical reasoning under bounded autonomy.

Concrete deliverables include: (i) a reference architecture outline with explicit responsibilities and auditable control points; (ii) a phase-gated deployment pathway (Preparation → Prototype → Pilot → Integration → AMAC operation) with required evidence packs, decision gates, and rollback/suspension mechanisms; and (iii) enforceable socio-technical gate criteria, including Socio-Technical Readiness Levels (STRL), readiness metrics, and accountability mapping (RACI). The model is intentionally non-normative and does not encode clinical guidelines; it provides a minimal, auditable governance architecture designed to make large-scale clinical AI integration feasible, controllable, and regulation-compatible in complex oncology environments.

Introduction and Context

Large-scale deployment of AI in oncology increasingly depends not only on algorithmic accuracy, but on system-level safety, accountability, interoperability, and regulation-aware governance. In large oncology centers in EU, AI solutions are introduced into complex socio-technical environments that combine heterogeneous IT infrastructures, evolving clinical workflows, and strict regulatory constraints under the EU AI Act and the Medical Device Regulation (MDR). As a result, the central challenge is no longer how to design isolated AI models, but how to deploy, govern, and evolve integrated AI platforms in a controlled, auditable, and compliance-oriented manner.

This shift aligns with broader observations that the main obstacles to clinical AI adoption increasingly include organizational, workflow, governance, and safety constraints alongside purely technical performance considerations (Jiang et al. 2021).

Recent AI maturity models in healthcare highlight organizational readiness, governance dimensions, and compliance constraints (e.g., Filipovic et al. 2026). Building on this perspective, the present work shifts attention from assessing readiness to operationalizing governance in regulated clinical environments through an explicit reference architecture, decision gates, and auditable deployment pathways.

This paper proposes the Architecture for Medical AI Collaboration (AMAC) as a deployment- and governance-oriented reference model for such platforms. **AMAC is a reference architecture paradigm that establishes a stable system-level envelope for the coordinated deployment, clinical governance, and gate-controlled lifecycle evolution of collaborative, multi-agent**

AI components within complex oncology environments. It is understood as a structured combination of (i) a reference system architecture outline, (ii) a reference system deployment pathway with explicit decision gates, and (iii) a set of auditable pre-requisites, roles, and responsibilities. The model is grounded in nearly one year of practical, pre-deployment experience from the OnkoBot project, including the development of preparatory mock-ups and proof-of-concept prototypes for multiple subsystems. Importantly, these artifacts were created during a preparatory phase; no clinical studies or clinical deployments are reported in this Part I. Instead, the experience serves as an engineering and governance basis for generalizing architectural boundaries, deployment prerequisites, and operational responsibilities.

Accordingly, references to the EU AI Act, MDR, and related standards are used in a compliance-oriented sense: they motivate the design of controls, roles, gates, and auditable artifacts, but do not constitute authoritative legal interpretation or a claim of regulatory conformity. In practice, compliance remains a validation-driven, site-specific outcome supported – rather than guaranteed – by the proposed mechanisms.

This paper was motivated by and abstracted from an extensive internal project charter developed jointly by NIO-PIB and UWM within a formal Letter of Intent (DĄBKOWSKI et al. 2025). The scope of this work is system-level deployment principles, safety-by-design mechanisms, and measurable operational indicators. Table 1 presents OnkoBot's main functional subsystems and the current status of work on mock-ups and prototypes.

Position within the three-part series. This article constitutes **Part I** of a three-part series. Part I establishes the foundational scope, definitions, architectural outline, and deployment pathway that are a *necessary precondition* for the more technical and formal developments addressed in Part II and Part III. In particular, Part II focuses on formal and algorithmic mechanisms for trust, evaluation, and decision gating, while Part III addresses extended validation, monitoring, and evolution under real-world operational constraints. Without the reference layer introduced here, such developments would lack a stable system-level context.

Key Scientific and Engineering Contributions:

Core contribution (Part I): Part I operationalizes EU AI Act/MDR constraints as a minimal, auditable online/offline governance contract – formalized as the AMAC reference architecture – and as a phase-gated deployment pathway applicable to integrated oncology AI platforms.

AMAC as an Auditable Governance Reference Architecture: Formal specification of the Architecture for Medical AI Collaboration (AMAC) as an implementation-neutral, system-level governance reference architecture that explicitly defines online/offline boundaries, auditable control points, and release conditions for clinical AI outputs.

Table 1
OnkoBot subsystem portfolio and proof-of-concept artifacts (illustrative, non-normative)

Subsystem	Primary purpose	Current PoC artifacts	Technical emphasis (interfaces / risk / governance)
OnkoBot.P	Patient/caregiver informational support	P1–P3 mock-ups/prototypes	Strict audience policies; safe templates; higher gating thresholds; provenance enforcement; HITL for high-risk queries.
OnkoBot.L	Clinician decision-support and workflow acceleration	L1–L4 mock-ups/prototypes	High-risk; interoperability dependence; OnkoTrust gating and HITL--first operation; traceable evidence.
OnkoBot.E	Education and adoption enablement	E1 prototypes	Sandbox and curriculum; controlled simulations; produces evaluation artifacts; supports safe usage patterns.
OnkoBot.B	R&D backbone for AI/KR methods	B1–B3 concept/prototype work	GraphRAG/KR pipelines; method evaluation; quantitative models; supports validated modules.
OnkoBot.K	Care coordination workflow support	Concept and early design work	Workflow integration; conservative policy-driven behavior due to operational impact.
OnkoBot.A	Audit, quality, and safety control	A1–A6 mock-ups/prototypes	Operational home of OnkoTrust: execution/auditing, regression tests, monitoring, incident workflows.
OnkoBot.D	Pathway analytics and organizational KPIs	Early planning work	Data pipelines and governance; aggregated analytics with strict interpretation constraints.

CEMA as a Bounded Multi-Agent Clinical Engine: Formalization of the Community of Collaborative Evolving Medical Assistants (CEMA) as a supervised, multi-agent computational core in which specialized agents collaborate under explicitly bounded autonomy, with all agent outputs released to end users only after centralized validation and governance gating.

Integrated Supervisory and Validation Stack: Definition of a centralized supervisory triad – OnkoTrust (symbolic grounding and internal consistency), QUANT services (statistical plausibility, uncertainty, and contradiction assessment), and the Quality Audit Agent (QAA) (offline governance, drift detection, and post hoc analysis) – providing a layered validation mechanism that is independent of individual agent implementations.

Regulation-to-Control Translation: Operational translation of EU AI Act and MDR requirements into concrete, enforceable engineering controls, including strict runtime immutability via online/offline separation, tamper-evident, append-only audit logging across system interactions, and mandatory human-in-the-loop (HITL) escalation paths bound to defined release and decision points.

Evidence-Driven, Phase-Gated Deployment Logic: Specification of a phase-gated deployment pathway (Preparation → Prototype → Pilot →

Integration → AMAC operation) in which progression is conditioned on predefined decision gates, mandatory evidence packs, and explicit rollback or suspension criteria, preventing uncontrolled transitions into higher-risk operational modes.

Socio-Technical Readiness as a Gate Condition: Introduction of Socio-Technical Readiness Levels (STRL) as measurable, enforceable gate-passage criteria that bind organizational preparedness, clinical workflow alignment, and governance maturity directly to deployment decisions, rather than treating readiness as informal background context.

Empirical Calibration from Pre-Deployment Engineering Practice: Calibration of auditable prerequisites, gate definitions, and evidence-pack structures using artifacts derived from approximately one year of pre-deployment engineering work within the OnkoBot project, including mock-ups and proof-of-concept systems, without making clinical or regulatory performance claims.

Clinical Governance Gateway (CGG) as a Release Authority: Specification of a dedicated Clinical Governance Gateway (CGG) that binds software releases, agent updates, and knowledge changes to documented clinical approval, safety verification, and accountability assignment, thereby separating technical evolution from clinical authorization.

Paper roadmap. Section (System Assumptions and Requirements for Large Oncology Centers) introduces system assumptions and requirements characteristic of large oncology centers. Section (Socio-Technical Readiness as a Deployment Prerequisite) addresses socio-technical challenges, organizational readiness, and human-in-the-loop aspects. Sections (Case-Guided Instantiation: OnkoBot) and (OnkoBot Reference Architecture Outline: The AMAC Framework) present the OnkoBot case-guided instantiation and the resulting reference architecture outline. Sections (Transferability to Smaller Centers) and (Reference Deployment Pathway) discuss transferability considerations and the reference deployment pathway. Finally, Sections (Discussion and Limitations)–(Further Research Directions) summarize limitations, conclusions with pointers to Parts II and III, and directions for further research.

For an alphabetically ordered list of abbreviations, see Table 2.

Table 2

List of abbreviations used in this interdisciplinary paper
(informatics, clinical oncology, governance, and regulation)

Abbreviation	Meaning / explanation
1	2
AI	Artificial Intelligence
AI Act	EU Artificial Intelligence Act: Regulation (EU) 2024/1689
AMAC	Architecture for Medical AI Collaboration (AMAC) is a reference architecture paradigm that establishes a stable system-level envelope for the coordinated deployment, clinical governance, and gate-controlled lifecycle evolution of interoperable, collaborative, multi-agent AI components within complex oncology environments

cont. Table 2

1	2
CEMA	Community of Collaborative Evolving Medical Assistants (intelligent multi-agent collective forming the clinical computational core of AMAC in clinical decision-support environments)
CER	Clinical Evaluation Report (MDR documentation artifact)
CGG	Clinical Governance Gateway (CGG) is a clinical governance function and architectural checkpoint that enforces evidence-based review and gate--controlled approval of AI component updates and data/knowledge releases. CGG enables auditable, criteria-driven release decisions based on predefined governance controls and verifiable evidence. If the available evidence is insufficient, inconclusive, or conflicting, CGG mandates a formal HITL escalation and grants authorization only upon documented clinical governance approval.
DICOM / DICOMweb	Digital Imaging and Communications in Medicine (imaging standard and web access)
EU	European Union
FHIR	Fast Healthcare Interoperability Resources (HL7 interoperability standard)
HIMSS	Healthcare Information and Management Systems Society.
GraphRAG	Graph Retrieval-Augmented Generation (RAG with graph-structured retrieval and provenance)
HIS	Hospital Information System
HITL	Human-in-the-Loop (formal human oversight workflow with auditable artifacts)
HL7	Health Level Seven (healthcare interoperability standards organization)
IAS	Identity, Access & Security: An identity-bound access control and security framework ensuring least privilege, accountability, and continuous auditability across the AI/IT ecosystem.
ID	Identifier (generic; e.g., patient, encounter, evidence)
IEC	International Electrotechnical Commission (standards body)
IEC 62304	Medical device software lifecycle processes standard
LIS	Laboratory Information System
LLM	Large Language Model
mCODE	minimal Common Oncology Data Elements (oncology data model on FHIR)
MDR	Medical Device Regulation: Regulation (EU) 2017/745
NIO-PIB	Maria Skłodowska-Curie National Research Institute of Oncology (Poland)
OnkoTrust	Trust layer concept (risk-aware gating, contradiction/grounding checks, escalation)
PACS	Picture Archiving and Communication System (imaging storage and retrieval)
PoC	Proof of Concept
QUANT	Quantitative/statistical consistency-check services (Quantitative & Statistical Gate)
RACI	Responsible, Accountable, Consulted, Informed (role assignment matrix)
RAG	Retrieval-Augmented Generation
RIS	Radiology Information System
RIS/PACS	Combined reference to radiology workflow system and imaging archive
RMF	Risk Management File (ISO 14971 documentation artifact)

cont. Table 2

1	2
SaMD	Software as a Medical Device (regulatory concept)
SBOM	Software Bill of Materials (software supply-chain documentation artifact)
SecOps	Security & Operations (technical enforcement function for security controls)
SIB	Secure Integration Bus: A secure integration layer mediating data and control flows across system components through policy-based routing, secure transport, and tamper-evident audit logging.
STRL	Socio-Technical Readiness Levels (maturity scale used for deployment gating)
UWM	University of Warmia and Mazury in Olsztyn (Poland)
XAI	Explainable AI (explainability methods / requirements)

Related Work

Research on artificial intelligence in healthcare spans a broad spectrum, ranging from algorithmic performance and explainability to ethical, legal, and organizational aspects of deployment. Early and influential surveys emphasize the importance of transparency, interpretability, and responsibility in AI systems, particularly in high-stakes domains such as medicine (BARREDO ARRIETA et al. 2020, HOLZINGER et al. 2019). These works establish conceptual foundations for trustworthy AI, yet largely remain at the level of principles, taxonomies, and design desiderata rather than operational deployment architectures.

A complementary stream of literature focuses on the ethical, legal, and liability implications of AI-assisted clinical decision-making. Studies by GERKE et al. (2020) and PRICE et al. (2019) highlight unresolved questions of responsibility, accountability, and risk allocation between clinicians, institutions, and technology providers. From a regulatory perspective, these concerns are formalized through binding legal frameworks such as the Medical Device Regulation (MDR) and the Artificial Intelligence Act, which impose strict requirements on lifecycle management, human oversight, traceability, and post-market surveillance for high-risk medical AI systems (European Parliament and Council 2017, 2024).

Another relevant body of work addresses interoperability and system integration in healthcare IT ecosystems. Standards such as HL7 FHIR and SMART on FHIR provide widely adopted mechanisms for secure data exchange and modular application integration (BENDER, SARTIPI 2013, MANDEL et al. 2016). While these standards are indispensable enablers of scalable AI integration, they do not by themselves define governance mechanisms, clinical decision gates, or accountability structures required for safe AI deployment.

More recently, maturity models for AI adoption in healthcare have been proposed to assess organizational readiness and critical success factors. Notably, FILIPOVIC et al. (2026) identify core dimensions such as strategy, data, governance, and skills as prerequisites for effective AI use in healthcare institutions. These models offer valuable high-level assessment frameworks; however, they stop short of specifying how AI components should be operationally governed, validated, and released within regulated clinical environments.

Parallel advances in multi-agent and collaborative AI systems demonstrate the growing technical feasibility of coordinated, evolving AI ecosystems (Chang 2025, Li et al. 2024, Air 2024). While these works illustrate the potential of collective intelligence and agent-based architecture, they typically assume research or experimental settings and do not address the regulatory, clinical governance, and risk-management constraints characteristic of large oncology centers.

Finally, granular computing and interactive computation frameworks provide a theoretical basis for structuring complex decision processes, uncertainty management, and explainable abstractions in intelligent systems (Pedrycz et al. 2008, Polkowski 2009, Jankowski 2017, Skowron et al. 2025). These foundations inform the design of auditable, threshold-based decision mechanisms but require explicit architectural embedding to support real-world clinical deployment.

In contrast to maturity assessment frameworks and principle-driven governance models, the present work focuses on a deployment-oriented reference architecture for medical AI. By integrating regulatory constraints, clinical governance checkpoints, gate-controlled lifecycle evolution, and auditable accountability mechanisms, the proposed approach aims to bridge the gap between conceptual readiness models and the practical realities of deploying AI systems in high-risk oncology environments.

System Assumptions and Requirements for Large Oncology Centers

This section specifies the system assumptions that underlie the proposed reference model. These assumptions are not presented as descriptive background, but as *explicit deployment prerequisites* that must be verified before advancing through successive stages of the deployment pathway introduced later in this paper. Failure to satisfy non-negotiable assumptions blocks progression beyond preparatory or pilot phases and requires corrective organizational or technical action.

Scope and hierarchy of assumptions. The reference model is intentionally scoped to *large oncology centers*, characterized by complex multi-specialty clinical workflows, heterogeneous IT infrastructures, and sustained regulatory oversight. Accordingly, assumptions are organized into two categories: (i) *non-negotiable prerequisites*, required for any compliance-oriented deployment of an integrated

AI platform, and (ii) *context-dependent assumptions*, which may be adapted based on institutional scale, maturity, and resource constraints. This distinction enables later transferability analysis without weakening baseline safety and governance requirements.

Non-negotiable organizational and governance prerequisites. At an organizational level, deployment assumes the existence of clearly assigned ownership for AI governance, including decision authority over model updates, deployment gates, and escalation procedures. Explicit roles for clinical experts, IT personnel, and compliance stakeholders must be defined, together with auditable processes for approval, documentation, and accountability. Human-in-the-loop (HITL) oversight is treated as a mandatory capability rather than an optional safeguard: qualified personnel must be available to review, override, or suspend AI-supported outputs whenever predefined conditions are met or exceeded.

In the reference setting, the oncology center is treated as the primary deployer of the integrated platform, while provider responsibilities for specific modules (e.g., AI services, monitoring, or integration components) may be assumed by an external vendor or the hospital IT unit, depending on the site’s governance and procurement model. This role split is intentionally left configurable, as it varies across deployments and determines the allocation of accountability and documentation duties.

Technical and interoperability requirements. From a system perspective, the reference model assumes a baseline level of IT interoperability and operational maturity. This includes stable interfaces for data exchange, explicit separation of offline training and evaluation environments from online clinical operation, version-controlled deployment and rollback mechanisms, and centralized logging that supports traceability and auditability. These requirements do not prescribe specific technologies, but define functional conditions that must be satisfied for safe integration into clinical workflows.

Interoperability Requirements and Operational Continuity. Interoperability is a first-order feasibility determinant for integrated AI systems in large oncology centers. In practice, such systems must interface with hospital information systems and electronic documentation modules (HIS/EDM), radiology information systems and imaging archives (RIS/PACS), laboratory information systems (LIS), and a variety of specialized oncology subsystems. These environments are typically heterogeneous and partially legacy. Consequently, interoperability should not be treated as an incidental integration task, but rather as a *dedicated subsystem* with explicit security boundaries, reliability mechanisms, and governance.

As a pragmatic baseline in typical European hospital IT landscapes, the interoperability layer often needs to handle HL7 v2/v3, FHIR, and DICOM/DICOMweb; the reference model remains implementation-neutral and does not mandate specific technologies.

Core functions include protocol translation, schema validation, policy enforcement (authentication, authorization, consent management, audit logging), and quality gates that prevent malformed or semantically inconsistent data from propagating into AI-supported workflows. Operational reliability mechanisms – such as bounded retries, dead-letter queues, and reconciliation jobs – are required to ensure predictable behavior under load and failure conditions.

Operational continuity further requires that the integrated AI platform degrades gracefully under partial failures. Temporary unavailability of upstream systems, delayed data feeds, or subsystem outages should not result in silent failure or undefined system behavior. End-to-end observability, including correlation identifiers across system boundaries, is assumed to be available to support auditing, incident response, and post hoc analysis of AI-assisted decisions.

Interoperability failure modes and mitigations are captured as auditable artifacts within the same gate-based deployment and release governance used across the platform (e.g., interface contracts, data-quality checks, incident runbooks, and integration test evidence). The resulting evidence packs consumed by governance review minimally include traceable log excerpts with correlation IDs (including SIB logs), integration and regression test reports, incident summaries (if any), and release-candidate configuration identifiers (version/hash) to enable reproducible audits.

Oncology-Specific Interoperability Profiles. Beyond generic HL7/FHIR and DICOM interfaces, oncology workflows benefit from domain-specific interoperability profiles that standardize data elements and clinical semantics across institutions. In particular, the mCODE initiative provides a structured oncology data model built on FHIR, enabling consistent representation of cancer diagnoses, staging, treatments, and outcomes. At the European level, the HL7 Europe FHIR Common Implementation Guide offers guidance on representing oncology concepts within FHIR-based exchanges.

The reference model assumes compatibility with such oncology-specific profiles where available. While local adaptations and extensions are often unavoidable, alignment with shared profiles improves portability, reduces integration friction, and supports secondary uses such as quality assessment and cross-institutional evaluation.

Regulatory framing as system requirements. Regulatory obligations under the EU AI Act and MDR are translated here into system-level requirements rather than legal claims. In particular, requirements for traceability motivate comprehensive logging and documentation artifacts; requirements for human oversight motivate explicit HITL roles and escalation paths; and lifecycle obligations motivate gated deployment, controlled change management, and post-deployment monitoring. The reference model is designed to support such compliance-oriented deployment, while recognizing that formal conformity assessment and clinical validation remain site-specific activities. Table 3 presents selected standards and regulations relevant to the proposed reference model.

Assumptions and deployment pathway integration. All assumptions introduced in this section are explicitly checked and enforced through decision gates in the reference deployment pathway presented in Section (Reference Deployment Pathway). Their role is therefore operational rather than descriptive: they determine whether a system may progress from preparatory work to pilot studies, integration, and compliance-oriented operation, or whether remediation is required before further deployment steps are permitted.

These assumptions are *regulation-informed but implementation-neutral*: they translate EU AI Act and MDR obligations – and the engineering expectations reflected in relevant ISO standards – into system-level prerequisites without prescribing particular technologies or organizational realizations. Detailed article-level mappings to the EU AI Act, MDR, and specific ISO clauses are

Table 3

Non-normative reference mapping of selected standards and regulations relevant to the proposed reference model

Standard / Regulation	Primary focus	Relevance to Part I
EU AI Act	Governance of high-risk AI systems, role separation (provider/deployer), and documentation and oversight obligations	Provides the governance framing for deployment-first design and accountability assumptions, without legal interpretation or conformity claims.
MDR (Regulation (EU) 2017/745)	Regulatory framework for medical devices and software as a medical device (SaMD)	Motivates lifecycle discipline, risk awareness, and documentation readiness for AI-supported medical software, without asserting device classification or compliance.
ISO 14971	Risk management for medical devices	Informs the identification of clinical risk hotspots, hazard analysis, and the linkage between risks and mitigation artifacts in the reference model.
IEC 62304	Software lifecycle processes for medical device software	Guides assumptions regarding controlled evolution, versioning, maintenance, and change management of AI assistants.
ISO 13485	Quality management systems for medical device organizations	Provides organizational context for roles, responsibilities, and documented processes, without implying certification or QMS implementation.
ISO 27001 / ISO 27799	Information security management and protection of health information	Supports assumptions related to secure interoperability, auditability, and operational continuity in integrated AI platforms.
GDPR (Regulation (EU) 2016/679)	Personal data protection, lawful processing, and data subject rights	Constrains data governance following GDPR, access control, consent/authorization practices, and auditability for patient-related data flows in CGG-enabled systems.

intentionally outside the scope of Part I and are addressed in later parts and supporting materials, once the foundational reference architecture and deployment pathway introduced here are fixed.

Socio-Technical Readiness as a Deployment Prerequisite

The deployment of AI systems in large-scale oncology centers is fundamentally a socio-technical transformation. Beyond algorithmic performance, the success and safety of the clinical mission depend on the alignment of human roles, organizational culture, and technical governance. This section operationalizes "organizational readiness" through quantifiable metrics and structured maturity levels, treating these factors as enforceable deployment prerequisites rather than contextual background.

The Centrality of Human Factors and Common Institutional Barriers

Even when AI models demonstrate high technical performance, systemic failures frequently arise from misaligned roles, opaque governance, or inadequate organizational readiness. In the context of Central and Eastern European (CEE) healthcare institutions, specific socio-technical barriers are particularly pronounced and can critically impede AI initiatives if not proactively managed:

- **Lack of Executive Sponsorship:** Insufficient "anchoring" of the project within the organization's top management, leading to resource constraints and strategic misalignment.
- **Motivation and Incentive Gaps:** Low engagement among clinical staff due to misaligned incentives, perceived threat to professional autonomy, or lack of visible benefit.
- **Communication and Silo Breakdowns:** Poor information flow and collaboration barriers between clinical, technical, administrative, and compliance departments.
- **Competency and Digital Literacy Gaps:** A misalignment between the required skills for AI-augmented workflows and the current capabilities of the workforce.

To navigate these complexities, this reference model adopts principles of socio-technical systems engineering. We refer to the Basic Principles of CSE Project Development (BPCD) as a non-normative but practical framework for governing the substantial organizational change inherent in AI-driven clinical transformation (Jankowski 2017).

Socio-Technical Readiness Levels (STRL)

To provide a structured, auditable path for organizational preparation, we introduce Socio-Technical Readiness Levels (STRL), inspired by established organizational maturity models in the CMU/SEI tradition. This scale ensures that the organizational environment matures in parallel with the technical infrastructure. Progress through the subsequent deployment pathway (Section (Reference Deployment Pathway)) is conditional upon reaching specific STRL milestones.

- **STRL 1 (Initial):** AI awareness exists at an individual level, but roles, responsibilities, and decision authority are ad-hoc and undocumented. No formal governance structure is in place.
- **STRL 2 (Defined): STRL 1 +** Governance ownership is formally assigned (e.g., a designated AI Steering Committee). Basic AI literacy and SaMD safety training programs for clinical staff are defined and implemented.
- **STRL 3 (Managed): STRL 2 +** Formal Human-in-the-Loop (HITL) roles and escalation paths are documented and verified through drills, establishing operational readiness without yet being exercised in live clinical decision-making. Interoperability protocols with key hospital systems (HIS/RIS) are established and operationally tested, **providing an ontological basis for a shared semantic and operational context across clinical departments, supporting consistent human communication as well as machine-to-system integration**.
- **STRL 4 (Predictable): STRL 3 +** Processes for monitoring and mitigating automation bias are active. Key Performance Indicators (KPIs) for AI safety, clinician burden, and system performance are regularly collected and reviewed by governance bodies (e.g., monthly), with documented thresholds and action triggers.
- **STRL 5 (Optimizing): STRL 4 +** The feedback loop is closed. Insights derived from CGG-controlled governance decisions, post-deployment monitoring evidence, and structured end-user feedback directly and systematically inform the iterative evolution of the platform, its workflows, and training programs.

Quantitative Readiness Metrics and the AI Ambassador Program

To support the EU AI Act requirements for human oversight (Art. 14) and institutional accountability, the reference model mandates the tracking of specific, quantifiable Socio-Technical KPIs. These metrics must be verified at each decision gate in the deployment pathway. Table 4 provides examples of such mandatory indicators. To actively mitigate the institutional barriers identified in Section

(The Centrality of Human Factors and Common Institutional Barriers), the model institutionalizes an **AI Ambassador Program**. This program designates respected clinical champions and operational facilitators who:

- Bridge communication between technical teams and clinical units.
- Lead peer-to-peer training and change management efforts.
- Gather and channel frontline feedback to the governance committee.
- Model safe and effective use of the AI system in daily practice.

Table 4

Exemplary Socio-Technical Readiness Metrics for Deployment Gate Review

Metric ID	Indicator	Threshold for Gate Passage	Rationale & Measurement Method
M-SOC-01	Stakeholder Alignment Index	> 85% positive engagement	Survey of clinical department heads regarding project goals, governance, and expected impact.
M-SOC-02	AI Literacy & Safety Certification	100% completion for HITL roles	Verifiable completion of mandatory training on SaMD fundamentals, limitations, and safety procedures.
M-SOC-03	Mean Escalation Response Time	< 5 minutes for high-risk triggers	Measured from system alert to clinician acknowledgment in the HITL interface during readiness drills.
M-SOC-04	Automation Bias Factor	< 0.15 (15% uncritical acceptance)	Rate of uncritical acceptance of seeded, simulated AI errors in controlled testing scenarios with clinical staff.
M-SOC-05	Audit Trail Completeness	100% of pilot interactions	Percentage of AI-assisted decisions in the pilot phase with a complete, retrievable log of input, context, evidence, and outcome.

Operationalizing Human-in-the-Loop (HITL) Oversight

Human oversight is operationalized not as a passive fail-safe but as an active, integral component of the workflow with defined triggers and artifacts. In clinical oncology, such HITL patterns are widely recognized as necessary to manage uncertainty, workflow risk, and safe escalation in real-world settings (YANG et al., 2022). The system architecture (see Section (OnkoBot Reference Architecture Outline: The AMAC Framework)) is designed to enforce HITL interception based on explicit Escalation Triggers. Table 5 defines these triggers and the corresponding auditable artifacts that must be generated.

The effectiveness of these triggers and the vigilance of HITL personnel are validated through periodic “Red Teaming” exercises, where synthetic failures and edge cases are introduced into the test system.

Table 5

Operational Governance Interface:
HITL Escalation Triggers and Auditable Artifacts

Trigger Category	Condition for Human Escalation	Auditable Artifact (Log)
Technical Uncertainty	Model confidence score below established threshold τ.	Evidence snapshot + raw model output.
Evidence Conflict	Discrepancy between RAG-retrieved clinical guidelines and LLM synthesis.	Conflict report + source document citations.
Safety/Risk Boundary	Detection of red-flag clinical indicators (e.g., life-threatening toxicity).	Full trace of safety-constraint violation.
Contestability	Manual override or ‚disagree' flag raised by the clinician.	Rationale for override + clinician ID
Ambiguity	Input data (e.g., pathology report) is corrupted or incomplete.	Data quality flag + missing field report

Accountability Mapping: The RACI Framework

Sustainable deployment requires unambiguous accountability. For every AI-supported workflow and output, a clear human agent must be accountable for the final clinical decision. This reference model adopts a RACI matrix (Responsible, Accountable, Consulted, Informed) to map accountability across all roles involved in AI-assisted care.

Critically, for any advisory output generated by the AMAC system, the **Accountable (A)** role is always assigned to a qualified clinical professional (e.g., the treating oncologist). The AI system and its operators may be **Responsible (R)** for generating the advice, but never **Accountable (A)** for the clinical outcome. This explicit mapping is a non-negotiable prerequisite for advancing from the Pilot to the Integration phase in the deployment pathway.

Integration with Architecture and Deployment Pathway

The socio-technical mechanisms specified here – STRL, metrics, Ambassador Program, HITL triggers, and RACI mapping – are not standalone recommendations. They are explicitly instantiated within the reference architecture (e.g., HITL triggers are enforced by OnkoTrust and QUANT Services) and are enforced as verification criteria at the decision gates of the reference deployment pathway (Section (Reference Deployment Pathway)). This integration ensures that organizational readiness is assessed with the same rigor as technical readiness before any progression to more advanced stages of clinical deployment.

Case-Guided Instantiation: OnkoBot

This section presents *OnkoBot* as a case-guided instantiation used to inform the proposed reference model. The purpose of this case is not to report a clinical deployment or clinical study, but to ground system-level design decisions in practical, pre-deployment experience. All OnkoBot elements discussed here correspond to preparatory mock-ups and proof-of-concept (PoC) prototypes developed during a preparatory phase; no clinical studies or clinical deployments are reported in this Part I. For orientation, we summarize the OnkoBot subsystem portfolio and representative mock-ups/prototypes developed across the program (Table 1). The table is illustrative and non-normative: it documents the decomposition used for engineering traceability and governance planning, without implying clinical readiness, regulatory classification, or deployment status.

Methodological role of the case. The case-guided approach adopted here serves to extract *system-level regularities* relevant to deployment and governance, rather than to generalize clinical outcomes. In particular, the OnkoBot experience is used to identify architectural boundaries, role allocation, auditable artifacts, and decision gates that recur across subsystems and use cases. This methodology is appropriate for constructing a reference model whose primary aim is to support controlled deployment under regulatory constraints, rather than to validate medical effectiveness.

Scope of preparatory work. Over nearly one year of preparatory work, multiple OnkoBot subsystems were explored through mock-ups and PoC prototypes, as summarized in Table 1. These artifacts were intentionally developed in a pre-deployment context to probe feasibility, governance implications, and integration challenges. They do not constitute medical devices, nor do they provide evidence of clinical effectiveness. Their role in this paper is illustrative and non-normative: they function as engineering probes that expose constraints and dependencies relevant to system-level design.

Extracted system-level lessons. The preparatory OnkoBot work yielded a set of recurring design insights that directly inform the reference model developed in this paper, including:

- the necessity of clear separation between offline training and evaluation environments and online clinical operation;
- the central role of auditable logging, documentation, and traceability across subsystem boundaries;
- the need for explicit human-in-the-loop (HITL) gating and escalation mechanisms to manage uncertainty and operational risk;
- the importance of clearly assigned decision authority for deployment, rollback, and exception handling;
- the operational relevance of continuous evaluation and monitoring activities beyond initial deployment.

OnkoBot as a narrative benchmark. Within this work, OnkoBot is treated as a *narrative benchmark* and design probe rather than as a reference implementation. Its value lies in anchoring abstract governance and deployment concepts in concrete preparatory experience, thereby reducing ambiguity when generalizing toward a reference architecture and deployment pathway applicable to large oncology centers.

Transition to the reference architecture. The observations and lessons summarized in this section directly inform the reference architecture outline introduced in Section (OnkoBot Reference Architecture Outline: The AMAC Framework). In the next section, these experience-grounded insights are consolidated into a structured architectural view that abstracts from individual prototypes while preserving the system-level constraints identified during the OnkoBot preparatory phase.

OnkoBot Reference Architecture Outline: The AMAC Framework

This section presents the core architectural contribution of this work: the Architecture for Medical AI Collaboration (AMAC) reference architecture. AMAC is a multi-agent, governance-first framework designed to enable the safe, compliant deployment of integrated AI platforms in oncology. Its design is explicitly shaped by two constraints: (1) lessons from the OnkoBot preparatory phase, and (2) the non-negotiable requirements of the EU AI Act and MDR for safety, predictability, and auditability.

OnkoBot Architecture visualization (case-guided illustration). To provide a comprehensive top-level overview, the OnkoBot architecture is presented through two complementary perspectives:

- **User-Oriented Architecture** – focuses on the functional aspects and how the system meets the needs of patients and clinicians. This perspective is analyzed from two distinct angles:
  - **User Journeys:** Mapping the end-to-end experience and interaction paths for both patients and medical professionals, as illustrated in **Figure 1**. The hospital IT/AI user-oriented architecture places a strong emphasis on the **comprehensive patient journey**, spanning from **prehabilitation** (preparation for treatment), through the **hospitaliza-tion** phase, to long-term **rehabilitation** and post-clinical follow-up.
  - **Functional Packages:** Categorizing the system's capabilities into logical modules of user-facing features, which are detailed in the **Core User Subsystems Table 1**.

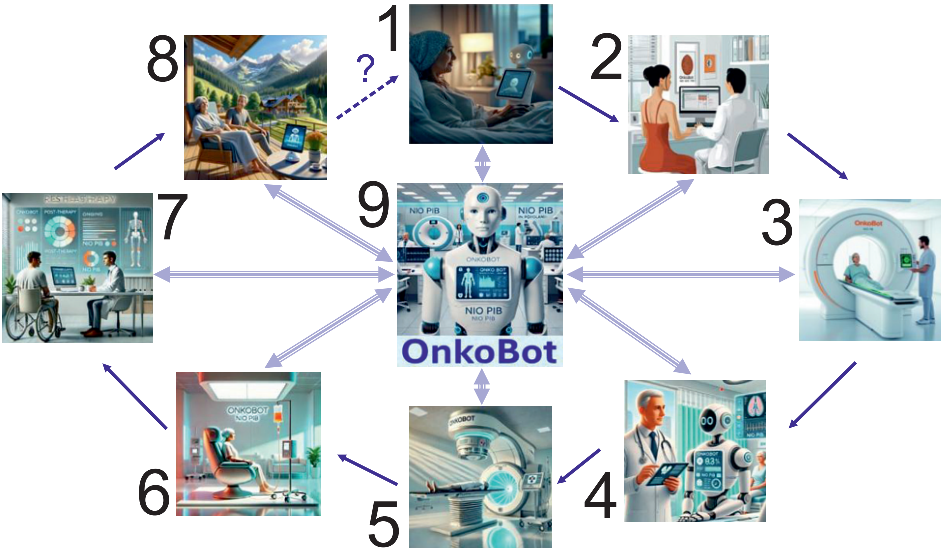


Fig. 1. OnkoBot closed-loop oncology pathway (illustrative, nodes 1–9). The diagram emphasizes unavoidable interactions among patients, clinicians, laboratories, medical equipment, and knowledge resources. Nodes (1–8) represent an illustrative patient journey from home support and consultation through diagnostics, therapy planning and delivery, and recovery support. Node (9) denotes the central orchestration core coordinating information flow and governance across stages. The dashed return arrow indicates the relapse/suspected--recurrence loop routing the case back to verification under governance control. The figure is a non-normative map of risk and validation focus rather than an exhaustive clinical taxonomy or a complete IT blueprint

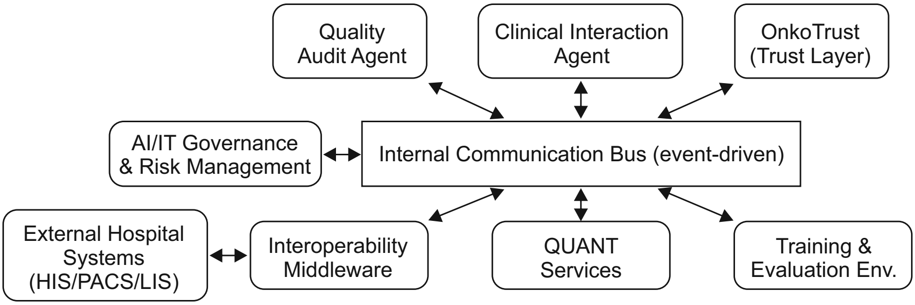

Quality Audit Agent
Clinical Interaction Agent
OnkoTrust (Trust Layer)
AI/IT Governance & Risk Management
Internal Communication Bus (event-driven)
External Hospital Systems (HIS/PACS/LIS)
Interoperability Middleware
QUANT Services
Training & Evaluation Env.

Fig. 2. Minimal system-level decomposition of the OnkoBot reference architecture. Co-operating agents and services are connected through an event-driven internal communication bus and bounded interfaces to external hospital systems (HIS/RIS/PACS/LIS and specialized subsystems). The AMAC community is governed via explicit trust and quality gates, including decision-time supervision (OnkoTrust and QUANT Agent/Services) and offline governance in the Training & Evaluation Environment (Quality Audit Agent). This online/offline separation supports auditable decision boundaries and controlled evolution through versioned releases rather than online self-modification during clinical operation

- **System-Level Architecture** – details the technical framework, including component interactions, data processing, and infrastructure requirements. This perspective emphasizes the underlying functionalities that ensure the reliable operation of the user-facing features. For OnkoBot, these top-level system functions and their dependencies are visualized in Figure 2.

From OnkoBot Experience to Generalized Architecture

The AMAC framework generalizes system-level insights gained from developing the OnkoBot portfolio of mock-ups and proof-of-concept prototypes (summarized in Table 1). Key design decisions in AMAC are direct responses to challenges encountered during this preparatory work:

- The need for strict role separation emerged from prototyping both patient-facing (OnkoBot.P) and clinician-facing (OnkoBot.L) subsystems, where failure modes and risk profiles differed significantly.
- The central importance of auditability was crystallized during the development of the OnkoBot.A (Audit) subsystem, which necessitated comprehensive logging and traceability across all components.
- The requirement for explicit, gated human oversight (HITL) was informed by early testing where ambiguous outputs required clear escalation paths to clinical experts.

Thus, AMAC does not describe a specific implementation but provides an implementation-neutral blueprint that distills these practical lessons into a reusable reference model for large oncology centers.

Architectural Overview and Core Principles

As noted above, the AMAC framework is visually summarized through two complementary perspectives that connect the clinical mission with technical execution.

The Clinical Pathway Perspective

Figure 1 illustrates the closed-loop oncology pathway that AMAC is designed to support. It maps the integrated patient journey from prehabilitation through treatment to follow-up (Nodes 1–8), emphasizing the unavoidable interactions between patients, clinicians, data sources, and AI orchestration (Node 9). This figure is not an exhaustive clinical protocol but a map of risk and validation focus. It identifies where in the patient journey specific AI functions (e.g., decision support in Node 4, therapy monitoring in Nodes 5–6) are deployed and, consequently, where architectural safeguards (like OnkoTrust gates) and validation efforts must be concentrated. The dashed "relapse/recurrence" loop underscores the system's role in continuous, longitudinal care under governance control.

The System Architecture Perspective

Figure 2 provides a minimal system-level decomposition of the AMAC reference architecture. It translates the clinical deployment pathway into a technical blueprint built around four core reference-architecture principles:

A. **Strict Online/Offline Separation:** The Clinical Operational Environment (online, right side of Fig. 2) is fixed at runtime. All learning, tuning, and updates occur exclusively within the isolated Training & Evaluation Environment (offline, left side). This makes the deployed system a predictable, fixed-function component and supports MDR-aligned evidence traceability to a specific software version.

B. **Governance-by-Design:** Auditability and human oversight are engineered as first-class system capabilities. Centralized supervisory gates – OnkoTrust and QUANT Services – enforce policy checks and safety constraints before any AI output can influence patient care.

C. **Multi-Agent Collaboration with Centralized Supervision:** The clinical computational core of AMAC is the Community of Collaborative Evolving Medical Assistants (CEMA), operating within the AMAC governance envelope under CGG-controlled gated approval of releases. CEMA is conceived as a set of specialized AI agents orchestrated by a central Clinical Interaction Agent (Fig. 2), whose autonomy is explicitly bounded by the Governance & Risk Management framework. All agent outputs are routed through the centralized validation stack (Audit/QAA, OnkoTrust, and QUANT Services), ensuring that clinical advice is validated, traceable, and verifiable prior to release to end users. This design is inspired by CHANG'S (2025) principles of regulated collective intelligence, adapted here to the constraints of large oncology centers under EU AI Act and MDR expectations.

D. **Integrated Governance & Risk Management:** AMAC consolidates risk-based controls, security governance, and compliance-oriented oversight through an integrated governance module. It integrates Security & Operations (SecOps) as the enforcement layer, within which Identity, Access & Security (IAS) delivers identity-bound access control, accountability, and auditability. Operational interactions between clinical and IT subsystems are mediated by a Secure Integration Bus (SIB), which enforces identity-validated access, secure transport, and policy-based routing. To support high-risk clinical use, the SIB maintains tamper-evident, append-only event logs, providing a transparent audit trail for system interactions. The Clinical Governance Gateway (CGG) serves as the clinical sign-off authority. It ingests auditable evidence packs (including SIB-derived logs) to enforce gated approval at review and transition checkpoints and to trigger formal escalation pathways when required.

AMAC Component Decomposition and Responsibilities

Operational Plane Components (Online)

Clinical Interaction Agent (CIA): The primary interface orchestrator. It receives user queries, decomposes them, and coordinates workflows among specialized sub-agents (e.g., for retrieval, summarization). It is responsible for context management and final answer synthesis.

OnkoTrust (Trust & Consistency Gate): The core safety module performing symbolic and rule-based checks:

- **Grounding Verification:** Ensuring statements are traceable to retrieved sources (guide-lines, records).
- **Contradiction Detection:** Identifying logical conflicts within the output or against trusted knowledge.
- **Policy Enforcement:** Applying institutional rules (e.g., "escalate all off-label sugges-tions").
- **Escalation Triggering:** Blocking outputs that fail checks and routing them to HITL with a conflict report.

QUANT Services (Quantitative & Statistical Gate): Provides data-driven checks:

- **Confidence Scores:** Based on model certainty and retrieval quality.
- **Statistical Plausibility:** Comparing suggestions against population norms.
- **Data Completeness Flags:** Assessing if available data is sufficient for reliability.

Interoperability Layer: A dedicated subsystem handling secure, reliable connections to hospital IT (HIS, RIS/PACS, LIS), performing protocol translation, validation, and resilience management.

Governance & Evolution Plane Components (Offline)

Quality Audit Agent (QAA): The central offline governance module. It analyzes logs from the operational plane, conducts periodic audits using synthetic and real dialogue logs, identifies performance drift, and generates evidence packs for regulatory audits and CGG-controlled evidence-based reviews.

Simulation & Training Engine: A sandboxed environment for training and evaluating new versions of agents, knowledge graphs (GraphRAG), and prompts against comprehensive test suites and simulated clinical scenarios.

Release Governance Module: Manages the gated pipeline for promoting changes from offline to online. It enforces that all updates pass regression testing, safety validation, and formal approval.

A brief summary of the proposed responsibility allocation across the Online/ Offline separation in AMAC is provided in Table 6.

Table 6

Responsibility allocation across the Online/Offline separation in AMAC

Aspect	Operational Plane (Online)	Governance & Evolution Plane (Offline)
Primary Purpose	Execute clinical decision-support tasks in real-time.	Evolve system knowledge, models, and policies under controlled conditions.
Key Modules	Clinical Interaction Agent, OnkoTrust, QUANT Services.	Quality Audit Agent, Simulation & Training Engine.
Learning/Adaptation	Prohibited. All parameters, prompts, and knowledge graphs are frozen.	Permitted via controlled cycles. Includes updating GraphRAG, fine-tuning, prompt engineering.
Change Mechanism	Changes only via versioned, audited releases from the offline plane.	Managed via gated release pipeline with validation suites and approval workflows.
Output	Clinical recommendations with associated confidence and evidence.	New software versions, updated risk files, validation reports, training datasets.

The Controlled Evolution Cycle and Transition Gate

AMAC replaces risky "online learning" with a formalized, auditable Controlled Evolution Cycle. This cycle, governed by a strict Transition Gate (Table 7), ensures that system evolution is both safe and compliant.

1. **Offline Development:** New models or knowledge graphs are developed in isolation.

2. **Shadow Mode Validation:** The candidate system runs in parallel with the stable version, processing real historical cases. Its outputs are logged and compared but not shown to clinicians, providing a risk-free performance assessment.

3. **CGG-controlled Review & Authorization:** Validation evidence and release artifacts are subject to review through the Clinical Governance Gateway (CGG) against predefined success criteria (e.g., non-inferiority on safety

metrics). When automated checks are insufficient, inconclusive, or conflicting, CGG triggers a formal HITL escalation and issues authorization for clinical deployment and use only after documented clinical governance approval.

4. **Gated Deployment:** Upon approval, the new configuration is frozen, hashed, and deployed as a new immutable version. Rollback procedures are always maintained.

Table 7

Transition Gate Requirements for moving a new AMAC version from Offline to Online operation

Gate Checkpoint	Verification Activity	Auditable Output
Functional Non-Regression	Automated testing against a curated "Golden Dataset" of complex clinical scenarios.	Behavioral Stability Report with pass/fail metrics.
Safety & Rule Compliance	Formal verification of adherence to all OnkoTrust rules. Execution of adversarial "Red Team" tests.	Updated Risk Management File (RMF) annex. Safety Test Report.
Clinical Validation	Blinded expert review of the new version's reasoning on challenging clinical vignettes.	Clinical Evaluation Report (CER) Addendum.
Configuration Lock & Sign-off	Final freeze and cryptographic hashing of the software bundle. Formal sign-off by the accountable governance body.	Signed Release Certificate (vX.Y.Z). Software Bill of Materials (SBOM).

Positioning AMAC within the Regulatory and Research Landscape

AMAC offers a pragmatic synthesis of two trends:

The Research Trend toward Agentic AI: It embraces multi-agent collaboration and long-term system evolution (*Institute for AI Industry Research* 2024), (Li et al. 2024).

The Regulatory Imperative for Safety: It strictly bounds autonomy within a governance framework that enforces determinism, auditability, and human oversight, directly addressing EU AI Act (*European Parliament and Council* 2024) and MDR (*European Parliament and Council* 2017) requirements. By institutionalizing the separation of operation and evolution, and by mandating CGG-controlled, evidence-based transition gating with formal HITL escalation where required, AMAC provides a reference blueprint for the compliant deployment of high-risk, evolving AI systems in clinical environments.

The AMAC architecture forms the foundation for the formal trust mechanisms (Part II) and long-term monitoring strategies (Part III).

Offline Evolution Cycle

To support the predictability, repeatability, and safety expectations associated with high-risk AI systems under the EU AI Act and MDR, the AMAC reference model mandates strict online/offline separation and runtime immutability of the Clinical Operational Environment.

Operationally, this separation requires controlled pathways for transferring validated changes from the offline environment into the online clinical workflow without compromising determinism at runtime, implemented through:

The Knowledge Transfer Mechanism: Transition Gates The migration of an "evolved" version of the AMAC from the offline environment to the online clinical workflow is governed by a formal **Transition Gate**. Under the framework of IEC 62304, any modification to agent logic or knowledge representation is treated as a new software release, requiring revalidation (see Table 7).

Shadow Mode and Clinical Benchmarking As an additional safety layer, the reference model introduces a **"Shadow Mode" Deployment**. Before an evolved AMAC version is permitted to provide active advice to patients or clinicians, it must operate in parallel with the stable version. In this mode, the new version generates recommendations that are logged, auditable, and subject to CGG-controlled review (with HITL escalation where required), while remaining invisible to end users. Access to the active clinical interface is granted only after no observed safety incidents above predefined thresholds and meeting predefined safety and governance criteria.

Regulatory Justification This modular-deterministic approach ensures that while the system remains "agentic" in its internal orchestration, it remains a "fixed-function" medical device during its operational lifecycle. This design supports MDR-aligned evidence traceability to a specific, immutable software version and enables human oversight over a predictable operational configuration, consistent with the governance expectations of the EU AI Act.

Reference Deployment Pathway

This section introduces a reference deployment pathway that operationalizes the reference architecture outlined in Section (OnkoBot Reference Architecture Outline: The AMAC Framework). The pathway is designed to support controlled, compliance-oriented rollout of an integrated AI platform by structuring deployment into staged phases separated by explicit decision gates. Progression through the pathway is conditional and auditable: advancement is permitted only when predefined organizational, technical, and governance pre-requisites are satisfied.

Pathway rationale and scope. The deployment pathway reflects the central premise of this Part I: large-scale AI deployment in oncology is primarily a systems and governance challenge rather than a purely technical one. Accordingly, the pathway emphasizes readiness verification, accountability, and controlled change over speed of adoption. It does not prescribe specific timelines or technologies, but defines a sequence of phases and gates that must be respected regardless of local implementation choices.

Phase Model with Explicit Review Gates The proposed reference model treats the *deployment pathway* not merely as a project plan, but as the conceptual backbone for designing, evolving, and governing both the integrated AI platform (OnkoBot) and its constituent sub-systems. In particular, the entire system as well as each user-facing and governance-facing subsystem is conceptualized through a shared *phase model*:

Preparation ⇒ **Mock-up/Prototype** ⇒ **Pilot** ⇒ **Integration** ⇒ **AMAC**.

Successive versions of subsystems traverse this pathway as modular, versioned building blocks metaphorically, *"LEGO blocks"* – that are incrementally built, tested, validated, and integrated under explicit governance and release gates. The pathway therefore unifies system architecture, development methodology, and organizational change within a single deployment logic.

Hospital-scale AI deployment should proceed through explicit phases with controlled scope expansion and formal exit criteria. Each phase concludes with a review gate evaluating readiness across four dimensions: safety, quality, interoperability, and governance. Advancement is conditional rather than automatic.

- **Preparation** establishes scope boundaries, assigns roles and responsibilities, identifies high-risk contexts, and assesses data availability, interoperability constraints, and security baselines.
- **Mock-up/Prototype** validates interaction patterns and architectural assumptions in controlled environments, typically limited to non-clinical or low-risk scenarios.
- **Pilot** introduces supervised, real-context use with mandatory human--in-the-loop control, exercising interoperability and operational continuity mechanisms.
- **Integration** embeds AI assistants into routine workflows across departments while preserving the same trust, safety, and audit constraints.
- **AMAC operation** supports long-term use and evolution of agents under controlled, auditable release cycles and explicit online/offline separation. AMAC is a multi-layer, agent-oriented reference architecture.

Each phase ends with a formal review gate that determines whether the next phase may begin.

Iterative Development and the Modular "LEGO" Principle Within the deployment pathway, each functional subsystem is treated as an independent, versioned module. Subsystems can progress through phases at different speeds, depending on risk profile and organizational readiness, while remaining interoperable through shared platform services.

The modular "LEGO" principle yields several operational benefits: failures are localized rather than systemic, validation efforts are focused, and integration is driven by governance readiness rather than technical enthusiasm. Importantly, modularity applies not only to technical components, but also to organizational artifacts such as training materials, procedures, and audit documentation.

Change Management: Ambassadors, Training, and Adoption Metrics Sustainable AI deployment requires structured change management alongside technical development. The reference model therefore embeds organizational adoption mechanisms directly into the deployment pathway.

Key elements include designated clinical and organizational *ambassadors*, role-specific platform literacy and training programs, sandbox environments for safe experimentation, and feedback loops capturing adoption metrics and trust dynamics.

Decision gates and verification. Transitions between phases are governed by explicit decision gates that evaluate whether required prerequisites have been met. These include verification of system assumptions, availability of HITL capacity, completeness of logging and audit artifacts, and readiness of escalation and rollback mechanisms. Decision outcomes are documented and traceable, ensuring that progression through the pathway produces auditable evidence rather than implicit acceptance.

Offline – online separation and change control. Consistent with the architectural principles defined in Section (OnkoBot Reference Architecture Outline: The AMAC Framework), all model updates, parameter changes, and policy adjustments are performed exclusively in offline environments. Online operation is restricted to execution under fixed, versioned configurations. Changes are introduced into operation only through gated releases following successful offline evaluation and formal approval, preventing uncontrolled adaptation during clinical use.

Integration of HITL and CGG. HITL oversight and CGG-controlled governance are enforced across all phases of the deployment pathway. HITL interception points are specified prior to pilot operation and may be tightened or relaxed only through documented, auditable governance decisions. CGG maintains a continuous governance feedback loop based on monitoring signals, incident reviews, and performance observations, which informs offline updates and gate-controlled release decisions without directly modifying online clinical behavior.

Rollback, suspension, and controlled degradation. The pathway explicitly incorporates mechanisms for rollback, suspension, and controlled degradation of automation. Trigger conditions for these actions are defined in advance and linked to monitoring and HITL inputs. The ability to revert to earlier phases or reduced functionality is treated as a core safety requirement rather than as an exceptional failure mode.

Pathway as a governance instrument. Beyond its procedural role, the reference deployment pathway functions as a governance instrument. It structures accountability, documents decision authority, and generates a traceable history of system evolution. In this way, it complements the reference architecture by ensuring that technical components, organizational roles, and regulatory expectations are aligned throughout the system lifecycle.

Transferability to Smaller Centers

This section addresses the transferability of the proposed reference model to smaller oncology centers. Transferability is not treated as free-form simplification, but as a controlled relaxation of assumptions defined in Section (System Assumptions and Requirements for Large Oncology Centers), performed under explicit constraints on safety, governance, and auditability. The objective is to preserve a non-negotiable core while permitting context-aware adaptation of scale-dependent elements.

Non-negotiable core (STRL ≥ 4 aligned). Independent of institutional size, the following elements are mandatory and must remain unchanged for any clinical deployment at STRL ≥ 4 (“Predictable”):

- minimum organizational readiness at STRL ≥ 4 (“Predictable”), ensuring documented, repeatable operational processes, deterministic online behavior, and auditable change control;
- mandatory adherence to the Reference Deployment Pathway, including phase-gated progression with explicit decision gates for deployment, rollback, and escalation;
- explicit assignment of governance ownership and decision authority for deployment, rollback, and escalation;
- enforceable, runtime human-in-the-loop (HITL) oversight with the ability to suspend or override automation;
- auditable logging, traceability, and version control across the system lifecycle;
- gated change management separating offline updates from online operation, with no ungoverned modifications in the clinical runtime environment;
- continuous evaluation and monitoring activities as an operational governance loop, including documented triggers for escalation, rollback, and release suspension.

Relaxation of these elements is not permitted, as it would undermine system-level safety and accountability. Transferability is therefore achieved exclusively by scaling the remaining components and organizational arrangements while preserving the non-negotiable core.

Scalable and adaptable elements. Other aspects of the reference model may be adapted to reflect reduced scale or resource availability. These include the depth of system integration, the number of automated components, the granularity of monitoring, and the organizational distribution of roles. Such adaptations are permitted provided that they do not weaken the non-negotiable core and remain verifiable through auditable artifacts.

HITL under resource constraints. In smaller centers, HITL capabilities need not be locally replicated in full. The model permits federated, shared, or centralized arrangements, including cross-institutional expert pools or external service models, provided that escalation paths, response times, and decision authority remain clearly defined and auditable. In all cases, insufficient HITL capacity constitutes a blocking condition for increased automation.

Risk–cost–complexity trade-offs. Transferability entails explicit trade-offs along axes of cost, automation level, HITL workload, and audit coverage, while maintaining a fixed clinical risk budget. Here, a “fixed clinical risk budget” denotes institutionally approved safety thresholds and escalation policies that are not relaxed when capacity is reduced; instead, automation and gating strictness are adjusted. Reductions in local capacity must therefore be compensated by more conservative automation, stronger gating, or shared governance arrangements, rather than by relaxing safety or oversight requirements.

Architectural and pathway implications. The reference architecture outlined in Section (OnkoBot Reference Architecture Outline: The AMAC Framework) supports modular scaling, allowing components to be included, simplified, or externally provided without violating core constraints. Likewise, the reference deployment pathway presented in Section (Reference Deployment Pathway) remains applicable across institutional scales, although smaller centers may require longer preparatory phases and more conservative progression through deployment gates.

Discussion and Limitations

This section discusses the scope, strengths, and limitations of the proposed reference model, with particular emphasis on its intended role as a deployment- and governance-oriented foundation rather than a clinical or regulatory validation study.

Scope and intended use. The reference model introduced in this Part I is designed to support controlled deployment, governance, and evolution of integrated AI platforms in large oncology centers. Its primary contribution lies in structuring architectural boundaries, organizational responsibilities, and deployment decision gates under regulatory constraints. Accordingly, the model targets system-level safety, accountability, and auditability, rather than algorithmic novelty or optimization of clinical performance.

Non-claims and deliberate exclusions. Several aspects are intentionally outside the scope of this work. First, this Part I does not establish clinical effectiveness, diagnostic accuracy, or therapeutic benefit of any AI component. Second, it does not by itself demonstrate regulatory compliance under the EU AI Act or MDR, as such compliance requires site-specific implementation, formal conformity assessment, and documented validation procedures. Third, detailed algorithmic specifications, parameter choices, and mathematical formalizations are deferred to subsequent parts of this series. These exclusions are deliberate and reflect a separation of concerns necessary for rigorous system design.

Experience-grounded but non-clinical basis. The reference architecture and deployment pathway are grounded in nearly one year of pre-deployment experience from the OnkoBot project, including the development of preparatory mock-ups and proof-of-concept artifacts. While this experience provides valuable insight into system-level constraints and governance challenges, it does not substitute for clinical studies or post-market surveillance. The model should therefore be understood as experience-informed rather than empirically validated in clinical practice.

Generalizability and context dependence. Although the reference model is intended to be applicable across large oncology centers, its instantiation necessarily depends on local context, including organizational maturity, IT infrastructure, staffing, and regulatory environment. Transferability to smaller centers requires controlled relaxation of assumptions, as discussed in Section (Transferability to Smaller Centers), and may involve federated or shared governance arrangements. Consequently, the model provides a structured framework for adaptation rather than a one-size-fits-all solution.

Implications for subsequent parts. The limitations identified here directly motivate the structure of Parts II and III. Formal trust mechanisms, evaluation criteria, and decision gating logic are addressed in Part II, while extended validation, monitoring strategies, and lifecycle evolution under operational conditions are explored in Part III. Together, these parts aim to complement the reference layer established in this work without overloading Part I with premature formal or clinical claims.

Architectural interpretation and evolution perspective. The proposed reference model deliberately separates architectural stability from the accumulation of system intelligence. AMAC defines a stable deployment and governance envelope whose role is to enforce controlled operation, accountability, and traceability, rather than to evolve into an autonomous clinical system. Within this envelope, clinical intelligence is progressively accumulated within the CEMA multi-agent core through controlled, offline updates and increasing agent collaboration. Crucially, lifecycle evolution never bypasses clinical governance: all clinically relevant outputs and releases remain subject to CGG-controlled, evidence-based review and gate-controlled authorization, with formal HITL escalation when automated checks are insufficient, inconclusive, or conflicting. In this sense, the Clinical Governance Gateway (CGG) functions as the permanent clinical governance function and architectural checkpoint governing all CEMA-driven outputs within AMAC, ensuring that increasing system intelligence remains bounded by invariant safety, accountability, and regulatory constraints.

Conclusions

The Architecture for Medical AI Collaboration (AMAC) defines an enforceable, system-level governance and deployment framework for integrated AI platforms in large oncology centers. It reframes the deployment challenge from isolated algorithmic performance to auditable architectural controls that constrain autonomy, regulate change, and operationalize safety, accountability, and interoperability requirements aligned with the EU AI Act and the Medical Device Regulation (MDR).

At the computational level, AMAC formalizes the Community of Collaborative Evolving Medical Assistants (CEMA) as the core clinical AI engine. CEMA instantiates a supervised, multi-agent architecture in which specialized AI agents perform coordinated clinical reasoning under explicitly bounded autonomy. This design deliberately aligns with contemporary multi-LLM collaborative intelligence paradigms – such as those articulated by Chang (2025) – while translating them into a governance-controlled clinical setting in which all agent outputs are subject to centralized trust, consistency, and release gating.

From a system perspective, AMAC establishes the following enforceable control principles:

(i) Runtime immutability and controlled evolution, achieved through strict online/offline separation that preserves deterministic clinical operation while confining model and knowledge evolution to gated offline environments;

(ii) Governed multi-agent operation, in which CEMA-generated outputs are admissible for clinical use only after passing centralized trust-and-consistency controls and explicit human-in-the-loop escalation paths;

(iii) Release accountability and traceability, implemented through a Clinical Governance Gateway (CGG) that binds deployment decisions to evidence packs, auditable approvals, and clearly assigned clinical responsibility;

(iv) Gate-based deployment conformance, preventing uncontrolled transitions between operational modes by enforcing phase-specific entry, rollback, and suspension criteria; and

(v) Socio-technical readiness as a mandatory condition, operationalized via Socio-Technical Readiness Levels (STRL), readiness metrics, and accountability mappings as formal gate-passage requirements.

Importantly, AMAC positions regulated multi-agent clinical systems as a realistic next step beyond single-model decision support. In this respect, it is consistent with emerging "AI hospital" initiatives – such as large-scale virtual hospital environments developed in China – while making a critical design choice: multi-agent clinical intelligence is acceptable only when embedded within a governance envelope that enforces auditability, bounded autonomy, and escalation-safe clinical accountability.

In summary, AMAC provides the system-level envelope required to make large-scale clinical AI integration feasible, controllable, and regulation-compatible in complex oncology environments. As Part I of this series, the present work fixes the architectural, governance, and deployment baseline. Subsequent parts build on this foundation by formalizing trust evaluation mechanisms, quantitative monitoring, and lifecycle feedback controls necessary for sustained, long-term clinical operation.

Further Research Directions

The reference model established in this Part I defines a stable system-level foundation for the deployment and governance of integrated AI platforms in oncology. Several directions for further research naturally follow from the scope delimitations and limitations discussed earlier, and are essential for completing the proposed framework across technical, formal, and operational dimensions.

Formal trust, evaluation, and decision gating. A primary direction for further research concerns the formalization of trust, evaluation, and decision gating mechanisms within the reference architecture. This includes the development of quantitative and logical models for confidence estimation, abstention, escalation, and acceptance under uncertainty, as well as their integration with human-in-the-loop oversight. Such mechanisms are addressed in Part II, where algorithmic and formal tools are introduced to operationalize these concepts without weakening governance constraints.

Advanced mathematical and statistical modeling. Further work is required to support rigorous analysis of robustness, calibration, and sensitivity across heterogeneous clinical contexts. This includes advanced mathematical and statistical modeling for uncertainty propagation, drift detection, and stress testing under varying data distributions and operational conditions. These methods are critical for moving from experience-grounded assumptions to quantitatively supported deployment decisions.

Extended validation and lifecycle governance. Beyond initial deployment, further research must address long-term validation and lifecycle governance of integrated AI platforms. This includes post-deployment monitoring, incident analysis, model update strategies, and mechanisms for managing concept drift and emerging risks under regulatory oversight. These topics are the focus of Part III, which examines how the reference architecture and deployment pathway can sustain safe operation over extended time horizons.

Cross-institutional and federated deployment models. Finally, additional investigation is needed into cross-institutional and federated deployment scenarios, particularly for smaller oncology centers. Such models raise new challenges related to shared governance, distributed HITL and AMAC functions, and coordinated auditability across organizational boundaries. Addressing these challenges is essential for scaling the proposed reference model beyond single institutions while maintaining safety and accountability.

Further research directions (deployment-first). Several issues merit further research: (i) systematic multi-center transfer studies with explicit capacity planning for HITL workloads and audit coverage; (ii) Interactive Granular Computing (IGrC) mechanisms for auditable, human-guided evolution of granules, thresholds, and operational policies over time (PEDRYCZ et al. 2008, POLKOWSKI 2009, SKOWRON et al. 2025); (iii) advanced mathematical modeling for quantitative robustness, calibration, and heterogeneity analyses across cohorts and clinical practice patterns (e.g., uncertainty calibration, shift/transfer diagnostics, and pre-defined statistical acceptance criteria for model updates); (iv) standardized *psycho-oncological* quality auditing protocols (synthetic and real-world) and their integration into post-market surveillance; and (v) long-term monitoring of drift, security threats, and governance effectiveness under evolving EU AI Act/MDR guidance.

Roadmap for future research on IGrC. The reference model motivates a transition from static granular representations toward interactive granular computing (IGrC), enabling auditable, human-guided evolution of knowledge granules, thresholds, and policies over time. Research in this direction aims to preserve traceability and control while allowing structured adaptation in response to new evidence, changing guidelines, or evolving organizational constraints. We plan to link the modeling of the AI systems discussed in the paper to the IGrC (JANKOWSKI 2017, SKOWRON et al. 2025). For more information, see https://dblp.uni-trier.de/pers/hd/s/Skowron:Andrzej. This will enable us to design and analyze AI systems based on the solid computational foundation of the IGrC and consider interactive granular computations over abstract and physical objects. The IGrC model can facilitate a more general approach than LLMs have thus far employed. For instance, it could enable us to examine the effectiveness of languages found in nature. Inspired by biology and other natural phenomena, these languages can advance reasoning tools for steering granular computations. This will also make AI systems more trustworthy and explainable (Barredo Arrieta et al. 2020) by providing explanations for suggested decisions, for example. Furthermore, applying the lifelong learning paradigm to AI systems will lead to continuous learning and the accumulation of past knowledge to assist with future learning and problem solving. This makes systems adaptable to new discoveries (e.g., outliers) and learning from past mistakes. One challenge of rough sets based on IGrC is developing high-quality classifiers that can determine whether information provided by LLMs is a hallucination and classify it with different degrees of risk accordingly. This will require advanced dialogue methods with domain experts. Another possibility is using IGrC to model c-granule control. This would make computational modeling of learning more similar to how the brain generates granular computations, constructing approximate solutions for given specifications.

Closing perspective. Together, these research directions delineate a coherent agenda that extends the foundational work presented in Part I. By progressively enriching the reference model with formal mechanisms, quantitative validation, and long-term governance strategies, future work can support the responsible and sustainable integration of AI into oncological practice.

References

Barredo Arrieta A., Díaz-Rodríguez N., Del Ser J., Bennetot A., Tabik S., Bar-Bado A., García S., Gil-López S., Molina D., Benjamins R., Chatila R., Her-Rera F. 2020. *Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI*. Information Fusion, 58: 82-115. Retrieved from https://www.sciencedirect.com/science/article/pii/S1566253519308103 (21.12.2025). https://doi.org/10.1016/j.inffus.2019.12.012

Bender D., Sartipi K. 2013. *HL7 FHIR: An Agile and RESTful approach to healthcare information exchange*. Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems (CBMS): 326-331. https://doi.org/10.1109/CBMS.2013.6627810

Chang E.Y. 2025. *Multi-LLM Agent Collaborative Intelligence: The Path to Artificial General Intelligence*. ACM Books. https://doi.org/10.1145/3749421

Dąbkowski M., Wawrzuta D., Żarłok E., Jankowski A., Polkowski L., Skowron A., Artiemjew P. 2025. *OnkoBot: Propozycja Karty Projektu. Projekt Zintegrowanego Systemu AI dla Narodowego Instytutu Onkologii PIB*. Internal project document (NIO-PIB and UWM), Warsaw/Olsztyn, version dated 5 October 2025. Available upon request from the authors (internal circulation).

European Parliament and Council. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation – GDPR). Official Journal of the European Union, L 119.

European Parliament and Council. 2017. Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices (MDR). Official Journal of the European Union. Retrieved from https://eur-lex.europa.eu/eli/reg/2017/ 745/oj/eng (21.12.2025).

European Parliament and Council. 2024. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union, L series, 2024/1689, 12 July 2024. Retrieved from http://data.europa.eu/eli/reg/ 2024/1689/oj (21.12.2025).

Filipovic K., Said R., Saban L., Uzdilli M., Krey M. 2026. *Toward an AI Maturity Model in Healthcare: Identifying Core Dimensions and Critical Success Factors*. Proceedings of the 59th Hawaii International Conference on System Sciences (HICSS-59): 6619-6633.

Gerke S., Minssen T., Cohen G. 2020. *Ethical and legal challenges of artificial intelligence-driven healthcare*. Artificial Intelligence in Healthcare: 295-336. https://doi.org/10.1016/B978-0-12-818438-7.00012-5

Holzinger A., Langs G., Denk H., Zatloukal K., Müller H. 2019. *Causability and explainability of artificial intelligence in medicine*. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 9(4): e1312. https://doi.org/10.1002/widm.1312

Institute for AI Industry Research (AIR), Tsinghua University. AIR. 2024. creates a virtual hospital, enabling AI doctors to self-evolve. Web publication, 24 May 2024. Retrieved from https://air.tsinghua.edu.cn/en/info/1007/1872.htm (21.12.2025).

International Electrotechnical Commission. 2015. *IEC 62304:2006+AMD1:2015 CSV – Medical device software – Software life cycle processes.

International Organization for Standardization. 2019. ISO 14971:2019 – Medical devices – Application of risk management to medical devices.

International Organization for Standardization. 2016. ISO 13485:2016 – Medical devices – Quality management systems – Requirements for regulatory purposes.

International Organization for Standardization. 2013. *ISO/IEC 27001:2013 – Information security management systems – Requirements.

International Organization for Standardization. 2016. *ISO 27799:2016 – Health informatics – Information security management in health using ISO/IEC 27002.

Jankowski A. 2017. *Interactive Granular Computations in Networks and Systems Engineering: A Practical Perspective*. Springer.

Jiang L., Wu Z., Xu X., Zhan Y., Jin X., Wang L., Qiu M. 2021. *Opportunities and challenges of artificial intelligence in the medical field: Current application, emerging problems, and problem-solving strategies*. Journal of International Medical Research, 49(3). https://doi.org/10.1177/03000605211000157

Li J., Lai Y., Li W., Ren J., Zhang M., Kang X., Wang S., Li P., Zhang Y.-Q., Ma W., Liu Y. 2024. *Agent Hospital: A Simulacrum of Hospital with Evolvable Medical Agents*. arXiv:2405.02957. Retrieved from https://arxiv.org/abs/2405.02957 (21.12.2025).

Mandel J.C., Kreda D.A., Mandl K.D., Kohane I.S., Ramoni R.B. 2016. *SMART on FHIR: A standards-based, interoperable apps platform for electronic health records*. Journal of the American Medical Informatics Association, 23(5): 899-908. https://doi.org/10.1093/jamia/ocv189

Pedrycz W., Skowron A., Kreinovich V. 2008. *Handbook of Granular Computing*. Wiley, Hoboken. Retrieved from https://onlinelibrary. wiley.com/doi/book/10.1002/9780470724163 (21.12.2025). https://doi.org/10.1002/9780470724163

Polkowski L. 2009. *Granulation of knowledge: similarity based approach in information and decision systems*. In: Encyclopedia of Complexity and Systems Science. Springer.

Price W.N., Gerke S., Cohen I.G. 2019. Potential liability for physicians using artificial intelligence. JAMA, 322(18): 1765-1766. https://doi.org/10.1001/jama.2019.15064

Skowron A., Jankowski A., Dutta S. 2025. *Interactive Granular Computing: Toward Computing Model for Complex Intelligent Systems*. Proceedings of the 20th Conference on Computer Science and Intelligence Systems (FedCSIS): 59-72. Retrieved from https://annals-csis.org/Volume_43/drp/pdf/6355.pdf (21.12.2025).

Yang J., Soltan A.A.S., Eyre D.W., Yang Y., Clifton D.A. 2022. *Human-in-the-loop machine learning in clinical oncology: challenges and opportunities*. The Lancet Digital Health, 4(10): e706-e708. https://doi.org/10.1016/S2589-7500(22)00145-3