AI & Future Society (2026), **2**(2), 43–67
DOI: https://doi.org/10.63802/afs.V2.I2.366

PANORAMA
SCHOLARLY GROUP

RESEARCH ARTICLE

Governing Artificial General Intelligence in Future Healthcare Institutions: Risk-Adaptive Institutional Authorization for Clinical and Operational Authority

Feifei Chen,† Chengwen Song,‡ Yuexin Lu,¶ and Qilong Wang†§

†School of Nursing and Health Management, Wuhan Donghu University, China
‡Department of Public Administration, Graduate School, Kookmin University, 77 Jeongneung-ro, Seongbuk-gu, Seoul 02707, Republic of Korea
¶Xi'an Medical University, 1 Xinwang Road, Weiyang District, Xi'an, Shaanxi 710021, China
§Gansu University of Traditional Chinese Medicine, Lanzhou, China
†Corresponding author. Email: Wade200023@outlook.com

Abstract

Increasingly general and agentic artificial intelligence raises an institutional question of authority: which actions may a system perform, how may it combine them, and who may revise or withdraw permission? For prospective artificial general intelligence (AGI), cross-domain planning makes unfamiliar combinations of familiar tasks a central problem of authorization. This article develops a Risk-Adaptive AGI Governance Framework through conceptual synthesis, purposive document comparison, and three constructed scenarios. Four determinants—task risk, demonstrated capability, institutional preparedness, and accountability—inform five authorization gates. The governing rule grants the least authority sufficient for a justified service objective within mandatory constraints. Information, recommendation, and bounded execution permissions can be combined only through additional authorization of their dependencies. A versioned record binds component permissions, configuration, and cross-service ownership; changes to a shared condition reopen affected grants. Administrative scheduling illustrates professional and patient challenges to an otherwise technically eligible grant. Emergency coordination traces partial execution, shared infrastructure failure, unavailable fallback, and conditional restoration. Four testable propositions address risk-sensitive allocation, preparedness, calibrated reliance, and precommitted containment. RAGF connects established lifecycle safeguards to continuing institutional authorization, dispute resolution, and patient remedy. The proposed evaluation agenda examines safety, workflow benefit, equitable access, and the costs of maintaining and withdrawing permission. For future healthcare institutions, preparation for AGI entails an accountable allocation of discretion across services, professionals, and patients.

Keywords: artificial general intelligence; healthcare governance; clinical autonomy; institutional preparedness; human–AI collaboration; patient safety; accountability

1. Introduction

Healthcare AI adoption redistributes authority as well as information. When a hospital introduces an AI tool, it changes who can access records, shape recommendations, initiate requests, and alter workflows. These actions are governed by professional jurisdictions, legal duties, and obligations

This article is published in *AI & Future Society* under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

to patients. A clinically promising system can fail when its outputs reach the wrong professional, recommendations escape review, or staff lack practical intervention powers. Governance therefore concerns which actions an institution permits, for which purposes and populations, and under whose authority.

Prospective artificial general intelligence (AGI) sharpens this allocation problem through cross-domain transfer and planning. A system could assemble clinical evidence, administrative resources, and departmental actions into a sequence beyond the workflows for which its components were evaluated. The resulting governance problem is how to authorize changing dependencies: permission for each component leaves open whether the new sequence has an acceptable purpose and consequences, and who is responsible for it. A task grant must therefore specify both the permitted action and its authorized connections to other tasks.

Existing healthcare AI guidance provides a foundation for this task. The World Health Organization (WHO, 2021, 2024) connects ethical principles with defined uses and oversight. The National Institute of Standards and Technology (NIST) AI Risk Management Framework links organizational governance to mapping, measurement, and management of risk (Tabassi, 2023). FUTURE-AI addresses trustworthy deployment across the lifecycle (Lekadir et al., 2025), while Joint Commission and Coalition for Health AI (CHAI) guidance addresses organizational implementation (Joint Commission & Coalition for Health AI, 2025).

Practice-oriented governance specifies institutional review as well as ethical principles. FAIR-AI combines risk-stratified review, a responsible business owner, an inventory and a Safe AI Plan (Wells et al., 2025). Freeman et al. (2026) connect tiered organizational oversight with structured review and address permissions, delegated authority, human approval, monitoring and shutdown for agentic applications. Continuing control also appears in NIST's deactivation provisions and applicable EU monitoring and suspension duties (European Parliament & Council of the European Union, 2024, Article 26; Tabassi, 2023, MANAGE 2.4). RAGF specifies how these controls remain connected across dependent action grants through shared eligibility conditions, coordinated restriction and accountable handover.

Recent agent research makes the allocation problem concrete. Zhang et al. (2026) combined on-premise operation with decision-time reliability signals and selective autonomy in retrospective diagnostic benchmarks. In those simulations, the signals informed case routing. Institutions must also determine which actions are permitted for a retained case, which professional is responsible for the decision, and which grants require review when patient context, configuration, or workflow changes.

This article asks: How should future healthcare institutions allocate and revise AI decision authority across tasks with different consequences and control requirements? It develops a Risk-Adaptive AGI Governance Framework (RAGF) in which autonomy is a conditional, revocable institutional authorization. Four dimensions structure the decision: task risk, capability evidence, organizational preparedness, and enforceable accountability. Five gates translate these dimensions into bounded permissions supported by documented reasons.

RAGF treats dependencies as conditions shared across grants. A change in patient eligibility, infrastructure, or workflow can require coordinated restriction and reassignment of work even while individual tools remain functional. Versioned authorization records allow these relationships to be inspected. The scenarios trace their consequences for professional jurisdiction, patient access, and responsibility for unfinished work. Four propositions translate selected mechanisms into expectations for future evaluation. The broader societal question concerns how institutions distribute discretion and preserve effective challenge as AI systems coordinate a growing share of care.

2. Research Approach and Scope

This study develops a conceptual synthesis through purposive comparison of governance documents and analysis of constructed scenarios. It examines how institutions allocate operational authority to increasingly general and agentic systems. AGI denotes a prospective capacity for transfer and planning across heterogeneous tasks; contemporary systems and governance documents provide evidence about the component practices through which such authority could be organized. The analysis draws on this evidence to derive authorization rules and examine their implications as combinations of work change.

Sociotechnical analysis locates health information technology within interdependent people, workflows, infrastructure, and external rules (Sittig & Singh, 2010). Automation research distinguishes functions and degrees of control and identifies mechanisms of overreliance (Goddard et al., 2012; Parasuraman et al., 2000). WHO (2021) supplies ethical grounds for human judgment, patient autonomy, and responsibility in health AI. Accountability identifies actors obliged to explain conduct to a forum able to scrutinize it (Bovens, 2007). The design also draws on least privilege in information protection (Saltzer & Schroeder, 1975).

The documentary corpus comprises WHO guidance (2021, 2024), NIST AI RMF 1.0, FUTURE-AI, DECIDE-AI, Joint Commission–CHAI guidance (2025), Regulation (EU) 2024/1689, and practice-oriented governance and implementation frameworks (Freeman et al., 2026; Wells et al., 2025; You et al., 2025). A. Wang et al. (2026) provide a scoping review of organizational governance frameworks; Hager et al. (2024), Goh et al. (2024) and Zhang et al. (2026) inform the distinction between task performance, clinician assistance and selective autonomy. The sources were selected for three functions: establishing permissible use, assessing context-specific evidence, and maintaining control over deployment. Normative documents supply principles and constraints; framework studies supply organizational procedures; empirical studies supply evidence about specified evaluation tasks.

The framework was developed in four steps. First, the selected documents and empirical studies were compared by governance unit, intended use, evidence requirements, decision ownership, and provisions for monitoring, intervention, and withdrawal. Second, their relevant mechanisms were grouped into four determinants of authorization: task risk, demonstrated capability, institutional preparedness, and accountability. Third, these determinants were translated into five sequential gates and a non-compensatory decision rule, with legal and ethical admissibility, adequate evidence, control readiness, and accountable ownership treated as conditions that each requested permission must satisfy. Finally, least privilege and complete mediation informed the scope and enforcement of the grant, while versioned records, shared-condition review, and capable handover were combined as the proposed institutional mechanism for managing dependent permissions. Table 1 and Section 2.1 distinguish inherited provisions from these design choices; the scenarios examine their implications under specific disturbances.

Table 1 compares these sources, and Section 2.1 traces four core rules to inherited principles, combined mechanisms, or explicit design choices. The scenarios examine how the rules operate together, including disputes among eligible options and failures that exhaust available controls.

Jurisdictional admissibility is a prerequisite to any institutional grant. Applicable law, professional scope, privacy and consent requirements, and product regulation determine which permissions a hospital can consider. The EU AI Act provides a legal comparator whose duties depend on system classification and the provisions applicable to a deployment (European Parliament & Council of the European Union, 2024). Institutional permission operates within those constraints.

Three constructed cases cover administrative scheduling, diagnostic support, and emergency coordination. They vary in reversibility, dependencies, and time pressure. Each specifies the requested permission, supporting evidence, responsible owner, disturbance, expected response, and failure criteria. An organizational dispute and a shared infrastructure failure test whether the rules yield a reasoned decision or identify an infeasible authorization. These analytical stress tests generate

implications for the empirical comparisons proposed in Section 4.6.

2.1 Positioning against existing governance frameworks

Existing frameworks address intended use, evidence, monitoring, and continuing oversight at complementary levels. WHO guidance establishes ethical and task-specific recommendations; FUTURE-AI specifies lifecycle practices; DECIDE-AI structures reporting of early live clinical evaluation (Lekadir et al., 2025; Vasey et al., 2022; WHO, 2021, 2024). Table 1 identifies their governance mechanisms and shows how RAGF translates them into local task grants. The derivation below identifies the source and institutional function of each proposed rule.

Table 1. Existing governance mechanisms and their translation into task-level authorization

Source and governance unit	Existing mechanism relevant to authority	Task-level implementation specified by RAGF
WHO (2021, 2024): Health AI and LMM governance	Ethical principles, well-defined tasks, stakeholder oversight, and post-release assessment.	Bind a local permission to its task, population, evidence, and review pathway.
NIST AI RMF (Tabassi, 2023): Organizational risk management	Govern, Map, Measure, Manage; MANAGE 2.4 addresses deactivation; GOVERN 1.7 and MANAGE 4.1 address safe withdrawal and recovery.	Attach the risk decision and stop mechanism to a versioned action grant.
FUTURE-AI (Lekadir et al., 2025): Trustworthy deployment	Local clinical validation; lifecycle risk management, logging, periodic audit, assigned oversight and risk mitigation including shutdown or alternative procedures.	Bind assurance evidence and assigned oversight to each action grant and its dependencies.
DECIDE-AI (Vasey et al., 2022): Early live clinical evaluation reporting	Reports clinical context, human factors, implementation, and safety in early evaluations.	Appraise reported early clinical evaluation evidence as one input to local permission decisions.
Joint Commission–CHAI (2025): Healthcare organization	Organizational governance, risk assessment, monitoring and staff education for responsible implementation.	Assign institutional roles for issuing, renewing, restricting and withdrawing local grants.
FAIR-AI (Wells et al., 2025): Institutional AI solution and intended use	Intake and business ownership; risk categorization; a Safe AI Plan with monitoring and transparency; a shared inventory.	Link solution-level review and monitoring to component and composition permissions affected by the same condition.
Freeman et al. (2026): Healthcare organization and proposed AI use	Tiered oversight matched to organizational maturity; structured review; agentic considerations include permissions, human approval, audit, runtime monitoring and shutdown or rollback.	Specify an assessed dependency range and propagate shared-condition changes to affected grants, outstanding work and capable receivers.
Clinical trials informed framework (You et al., 2025): Staged implementation	Safety, efficacy, comparative effectiveness and post-deployment monitoring, including workflow assessment and de-implementation.	Match the evidence and review interval for each grant to its proposed stage, task and recovery conditions.
EU AI Act (European Parliament & Council of the European Union, 2024): Relevant regulated systems and deployments	Risk management, competent human oversight, monitoring, and applicable suspension obligations (Articles 9, 14 and 26).	Record local responsibilities, action boundaries, and enforceable suspension within applicable law.
RAGF: Institutional task authorization	Author-specified conjunction and dependency linkage, applying least privilege and complete mediation.	Bind component grants to shared conditions, capable handover, and reviewable authorization reasons.

Note. The final column specifies the proposed institutional implementation of each source. Legal force and scope differ across sources.

A. Wang et al. (2026) identified 77 acute-care governance frameworks in a search updated through March 2025; 10 included guiding principles, an assessment method, lifecycle stages and an oversight mechanism. The FAIR-AI and Freeman frameworks provide concrete review and implementation procedures. These sources make institutional review, monitoring and withdrawal substantive starting points for the present synthesis. RAGF takes the continuing relationship among component permissions and their assessed composition as its design object: one shared-condition change can alter several grants and the allocation of unfinished work.

Four derivations make the synthesis traceable. First, WHO (2024, Section 5.2, p. 48) states that human rights and ethical obligations must be upheld irrespective of a system's risk or benefit. RAGF carries this constraint into authorization and extends conjunctive eligibility to evidence, control readiness, and accountable ownership as a normative design choice. Second, least privilege is inherited from Saltzer and Schroeder (1975, Section I.A.3.f); RAGF extends its task requirement to a service objective justified through professional and patient obligations. The extension determines which eligible authority is sufficient, with adequacy assessed separately from narrowness.

Third, complete mediation (Saltzer & Schroeder, 1975, Section I.A.3.c) supplies the principle of checking each access and refreshing changed authority; binding that check to configuration and clinical context is the proposed implementation. Fourth, NIST's safe decommissioning, assigned deactivation, and recovery mechanisms (Tabassi, 2023, GOVERN 1.7, MANAGE 2.4 and 4.1) support withdrawal. RAGF combines them with cross-service ownership to require reconciliation and capable handover for each affected grant. The conjunction, dependency linkage, and clinical handover duties are author-specified design relations; Table 1's source summaries provide their institutional context.

3. From Clinical AI to Institutional Decision Authority

3.1 General capability and task evidence

Predictive models, large multimodal models, and tool-using agents describe different system properties. Prediction concerns an estimated outcome; multimodality concerns inputs and outputs; agency concerns the capacity to initiate and sequence actions. WHO (2024) treats broad foundation-model capability as a matter requiring evidence. Healthcare institutions can therefore evaluate the specific task, setting, and failure modes relevant to permission without using an AGI label as a substitute for demonstrated competence.

Here, AGI denotes a prospective capacity to transfer knowledge, adapt to unfamiliar cognitive problems, and plan across organizational contexts with limited task-specific reengineering. This breadth can lower the effort needed to recombine tasks, allowing the reachable action sequence to change faster than a fixed workflow approval. RAGF therefore binds coordination to an approved dependency structure: new task relationships outside the assessed range require review even when every component retains valid evidence. General capability intensifies the need to assess how tasks are combined and how those combinations change.

A similar mechanism can arise with narrower contemporary agents. Consider separately approved summarization, appointment changes, and transport booking. Their combination could repeatedly move a patient along a pathway inferred from a summary. Permissions for the components can remain valid even as the combined sequence changes cumulative delay and responsibility. Under

RAGF, that new connection requires a coordination grant; a shared input correction can restrict every dependent execution grant. Section 5.1 develops this mechanism through scheduling and its dependent actions.

3.2 Hospitals as interdependent sociotechnical organizations

Hospital tasks depend on distributed expertise, incomplete information, and handovers. Rescheduling alters access to imaging, staffing, and continuity; diagnostic advice changes subsequent investigations. Sociotechnical analysis places technology within these relationships (Sittig & Singh, 2010). Assessment therefore includes the workflow through which an output affects care.

Cross-department agency creates a propagation mechanism. An incorrect assumption can be copied into a summary, used to prioritize a consultation, and then inform a scheduling action. Repetition can make the assumption appear independently corroborated. Risk depends on the actions the system can sequence, whether they occur concurrently, and the opportunity to interrupt them. Even validated tasks require additional workflow assurance when their combination correlates errors or makes recovery difficult.

Overreliance on automated advice can displace active checking and encourage acceptance of erroneous recommendations (Goddard et al., 2012). Sustained delegation of a cognitive function also creates a risk of skill disuse (Parasuraman et al., 2000, Section IV.A.4). RAGF therefore evaluates whether staff retain the evidence, attention and intervention powers needed for the authorized task, and whether recovery arrangements remain usable as services become dependent on automated workflows.

Song et al. (2026) theorize cognitive sacredness as a condition in which AI acquires elevated epistemic authority and deference becomes insensitive to verification. This perspective adds an authority-attribution mechanism to the analysis of reliance. In RAGF, staff and patient routes to question recommendations and seek review of the governing permissions provide practical means of contesting AI-supported decisions.

A randomized vignette study involving 50 physicians found no statistically significant improvement in diagnostic reasoning from access to GPT-4 plus conventional resources compared with conventional resources alone; an exploratory model-alone analysis yielded higher reasoning scores than the conventional-resource arm (Goh et al., 2024). Hager et al. (2024), using retrospective MIMIC-IV-derived cases, evaluated information gathering, guideline adherence and input sensitivity alongside diagnostic performance. These studies motivate separate assessment of model performance, clinician–AI collaboration and workflow consequences, with evidence matched to the task and setting.

3.3 Professional authorization and institutional accountability

Professional authorization identifies the qualified actor entitled to make or approve a consequential care decision within the applicable scope of practice. WHO (2021, Sections 5.1 and 5.4) connects human control and clinical judgment with responsibility and accountability. Bovens (2007) defines accountability through an actor's obligation to explain conduct to a forum that can question and judge it, with the actor potentially facing consequences. Together, these relationships connect delegation to legitimate decision makers and require reasons that can be reviewed. Capability evidence supports a grant; professional mandates and patients' rights constrain its scope.

Licensing, organizational and product duties, and patients' rights shape the exercise of authority in healthcare. A task grant connects these arrangements to control over clinical judgment, procurement, interfaces, maintenance, and monitoring. RAGF allocates prospective duties according to each actor's control over the relevant action and opportunity to prevent harm, within applicable professional and legal obligations.

RAGF assigns technical change duties to vendors, local authorization to institutions, operational supervision to services, and consequential care decisions to responsible professionals. An independent forum reviews the arrangement. Staff and patients can challenge both execution and its authorizing decision.

4. The Risk-Adaptive AGI Governance Framework

4.1 Foundational principle: autonomy is an institutional authorization

Autonomy is an institutional grant to perform specified actions under continuing conditions. Capacity concerns what a system demonstrates in relevant evaluations; permission identifies which of those activities the institution allows in a particular workflow. The grant binds a task to a population, data and tool boundaries, system configuration, responsible roles, and review period. Its scope can differ across services and be restricted as evidence or operational conditions change.

Four determinants guide the authorization decision. Task risk concerns credible harms, their likelihood and severity, reversibility, time pressure, patient exposure, and downstream propagation. Capability concerns task validity, subgroup performance, reliability under missing or conflicting inputs, and safe responses to uncertainty and tool failure. Preparedness concerns local data quality, staffing, interfaces, escalation, and fallback capacity. Accountability concerns a named authorizing authority, traceability, contestability, audit access, vendor duties, and an enforceable power to stop execution.

The four determinants define the assessment criteria, the five gates structure the decision, and a versioned permission record specifies the resulting authority. Legal and professional admissibility constrain every output. Adequate task evidence, feasible control, and accountable ownership are conjunctive prerequisites. Benefit and cost guide selection among eligible arrangements. Under the proposed non-compensatory rule, each prerequisite can veto the requested action; the hospital records the evidence and reasons for its decision.

4.2 The five authorization gates

Gate 1: Legal and ethical admissibility. The authorizing service identifies applicable law, professional scope, privacy and consent requirements, procurement obligations, and patient rights. The service records the jurisdiction and grounds for the proposed use. Patient-facing processes provide appropriate disclosure and a route for review. Unresolved admissibility questions keep the affected action outside the grant until competent clinical and legal authorities determine its status.

Gate 2: Task and dependency classification. The hospital identifies the requested action, intended benefit, affected population, connected systems, and worst credible harms. The classification includes downstream clinical dependencies and uncertainty in the evidence. It flags irreversible consequences, vulnerable populations, ambiguous clinical inputs, and time pressure. The resulting record sets a ceiling on execution rights and identifies the events requiring reassessment.

Gate 3: Capability and assurance evidence. Evaluation addresses the intended action in its proposed setting, including omissions, subgroup harms, deferral, and dependent systems. Retrospective validation, simulation, shadow operation, and prospective evaluation support different deployment decisions. FUTURE-AI provides lifecycle assurance practices; DECIDE-AI provides reporting expectations for early live clinical evaluations (Lekadir et al., 2025; Vasey et al., 2022). For patient-facing information support, assurance should assess information quality, source transparency, readability, and empathy separately, as illustrated by a comparative evaluation of eight LLMs answering public mental health questions (Q. Wang et al., 2026). The authorizing body documents how the evidence supports each requested permission.

Gate 4: Organizational control readiness. The service demonstrates who observes execution, who can intervene, how quickly the intervention reaches connected systems, and how care continues during interruption. Staff workload and availability are assessed under realistic conditions. Access

controls keep authority within approved boundaries. An independently controlled mechanism blocks unauthorized actions and enforces required deferral.

Gate 5: Conditional authorization and renewal. The authorized decision maker issues a versioned grant specifying tools, input boundaries, review requirements, monitoring indicators, expiry, and suspension triggers. A clinical service accepts operational ownership, and a governance forum accepts responsibility for review. Material changes in the model, prompts, retrieval sources, interfaces, population, or workflow reopen the relevant gates. Renewal records the evidence that supports continued use.

Recommendation and execution require distinct permissions. A diagnostic suggestion shapes a clinician's decision; a tool call can submit an order or change a booking. Assessment follows each pathway to its consequences. Execution requires enforced access limits, recovery arrangements, and explicit permission for the transaction through which an output takes effect.

4.3 A non-compensatory authorization rule

An authorization unit records the task, intended benefit, patient or operational population, permitted inputs and outputs, tool interfaces, versioned configuration, responsible roles, and review interval. Legal admissibility is checked first. The risk assessment then defines prohibited actions and an authority ceiling; evidence and organizational controls determine the feasible permissions within it. Accountability provisions identify who grants, supervises, and reviews the use. If a mandatory condition is unmet, the requested action remains unauthorized. Any narrower grant or evaluation configuration is assessed against its own applicable gates.

The decision rule is to grant the least authority sufficient to achieve a justified service objective within every mandatory boundary. Least privilege supplies the inherited selection principle (Saltzer & Schroeder, 1975); professional and patient obligations determine which objectives and constraints are legitimate. The committee compares eligible grants with narrower options and the existing workflow, documenting expected benefits, delays, burdens, and recovery costs. A narrower grant is sufficient only if it adequately serves that objective. Section 6.1 specifies how competing definitions of necessity are decided and reviewed.

4.3.1 Authorization records and enforceable transitions

A usable record contains a unique grant identifier; the task and population; permitted data, outputs, and tools; excluded actions; the model, prompt, retrieval, and interface configuration; linked evidence; references to shared conditions and their maintaining services; the authorizing role, operational owner, and reviewing forum; action and timing limits; monitoring signals; fallback procedures; and an expiry or review date. Grant states include proposed, active, restricted, suspended, expired, and withdrawn. Activation and reinstatement require recorded approval from the designated authority; previously approved rules or an authorized person can trigger restriction or suspension.

Before consequential execution, an independently controlled policy layer checks the grant, configuration, and authoritative context. The clinical service maintains eligibility and urgency; the agent has no write access to those fields or to its grant. The receiving interface must bind the check to a context version or repeat it when the transaction commits. A changed or unknown state blocks the affected transaction. This is a functional assurance requirement: an interface that cannot enforce the check supports drafting or submission by a human after prior review for that action. Section 5 traces blocked, partially executed, and completed work through containment and recovery.

Worked authorization record: nonurgent appointment rescheduling. The following completed record is a constructed example based on Scenario A. Its identifiers, review periods, controls, and evidence illustrate a local decision; a hospital would replace them with its own approved criteria and documented evidence.

Grant and status. Grant SCH-E-01, version 1.0: active following recorded approval by the hospital's authorized AI governance officer. The grant expires 90 days after activation unless renewed; the operational owner reviews it every 30 days and after a recorded trigger.
Purpose, population, and scope. Reduce avoidable scheduling conflicts for adults on a designated nonurgent outpatient pathway. The E grant permits the scheduling interface to move one appointment within the clinical window maintained by the pathway owner. A linked C grant, SCH-C-01, permits coordination of that appointment with its already approved transport booking; it supplies no additional execution rights. The example grant permits no more than one change per appointment. The patient receives a notification and a nondigital route to request review.
Inputs, tools, and exclusions. Permitted inputs are the appointment identifier, available slots, the current eligibility and urgency status, the clinically approved time window, and the linked transport booking. The agent may use only the approved scheduling and transport interfaces. It cannot modify urgency, clinical priority, orders, or the clinical window; move an urgent or ineligible appointment; or add a new service or dependency. The clinical pathway service maintains clinical eligibility; the agent cannot alter it.
Configuration and shared conditions. The grant applies to scheduling agent version 1.0, prompt version 1.0, an approved scheduling interface version 1.0, and an approved transport interface version 1.0; no external retrieval source is enabled. SCH-C-01 records the appointment–transport dependency. The clinical service maintains current eligibility; the scheduling and transport services maintain their respective interface availability and transaction records. The policy layer checks the active grant, configuration, eligibility, and context version at the point of execution. A changed or unknown context blocks the affected transaction and pauses dependent transport work.
Gate decision and evidence. G1: the hospital's clinical and privacy leads record that the specified use is admissible under applicable local requirements. G2: the pathway owner records the harm of delayed access and approves the nonurgent population and time window. G3: the assurance team links local simulation, subgroup access checks, and a shadow comparison with schedulers to the grant. G4: the operations lead links a successful commit-time check, a tested stop control, and a staffed manual fallback to the record. G5: the authorizer approves the E and linked C scopes after reviewing these gate records. These are fictional example entries, not reported hospital results.
Owners, monitoring, and triggers. The scheduling operations lead is responsible for execution, while the clinical pathway owner maintains eligibility; an independent safety or ethics forum reviews disputed authorization decisions. The operations lead monitors blocked transactions, repeated changes, delays, patient complaints, and unassigned work. An urgency change blocks the affected transaction; a failed policy check or unavailable trusted eligibility data suspends affected execution. Repeated disruption or a material change to configuration, population, or dependencies reopens the affected gates. An authorized staff member may restrict or suspend the grant.
Handover, remedy, and renewal. When execution stops, the scheduling service assigns pending work to a named human scheduler, reconciles any messages or transport actions already sent, and confirms that the receiving service can act. The pathway service reviews patient challenges, corrects records, and restores necessary access where feasible. Reinstatement requires restored controls and trusted context, reconciled work, available handover capacity, and recorded approval by the authorizer. Before expiry, the reviewer records the renewal decision and supporting evidence.

4.4 Forms of authorized action and coordination

RAGF distinguishes three action permissions: information support (I), recommendation (R), and bounded execution (E). An E grant names the transaction, trigger, recipients or objects, limits, and required approvals. Low-harm reversible actions are one eligible use; higher-consequence transactions require assurance and preventive controls appropriate to their own consequences. Where adequate control depends on a professional decision, that decision remains a prerequisite to execution. The

scenarios reserve diagnosis, prescribing, treatment, and changes to clinical orders to the responsible professionals. Risk is assessed for every form, including misleading information.

Coordination (C) is an additional permission over assessed task relationships and a specified range of permitted variation, combining any authorized I, R, or E components. A recommendation-only sequence can require C; its presence conveys no additional execution rights. Both components and the combined workflow must pass the gates. Individually low-risk actions can therefore form an ineligible combination. In emergency communication, E can permit transmission of a clinician-activated alert to specified recipients; C covers its escalation sequence. The clinical trigger and treatment decisions remain outside the agent's discretion. The C record specifies participating tasks and services, permitted dependencies, population and purpose, timing constraints, and responsible owners. Replanning within this range uses current-context checks and the grant's approved limits or review requirements. A transaction can lose eligibility under changed patient context while the assessed composition remains valid. A new dependency or an expansion beyond the assessed population, purpose, consequences or responsibility range reopens the affected gates before execution. Shared-condition changes also trigger the recorded reassessment or suspension rules for affected grants. Table 2 distinguishes action forms from composition.

Table 2. Proposed forms of institutional authorization

Permission	Permitted activity	Required control arrangement	Illustrative application
I: Information support	Retrieve, summarize, or draft within a specified information workflow.	Source access limits, provenance, and review before consequential use.	Prepare a handover summary for checking.
R: Recommendation	Present options and decision support to qualified staff.	Clinical review, visible uncertainty, and separate authorization of consequential actions.	Suggest a differential diagnosis.
E: Bounded execution	Perform named transactions with explicit triggers and boundaries.	Risk-specific evidence, trusted context checks, action limits, and required professional approval.	Eligible appointment changes; transmission of clinician-activated alerts.
C: Additional coordination permission	Compose approved I, R, or E actions within assessed task relationships and permitted variations.	Component and whole-workflow gates; dependency checks, ownership, escalation, and handover.	Link appointments and transport; route an approved alert sequence.

Note. I, R, and E describe action forms. C authorizes their composition. Component and whole-workflow eligibility are assessed separately.

Figure 1 links the five authorization gates to versioned permissions, execution checks and the review or recovery paths triggered by changes in context, scope or control conditions.

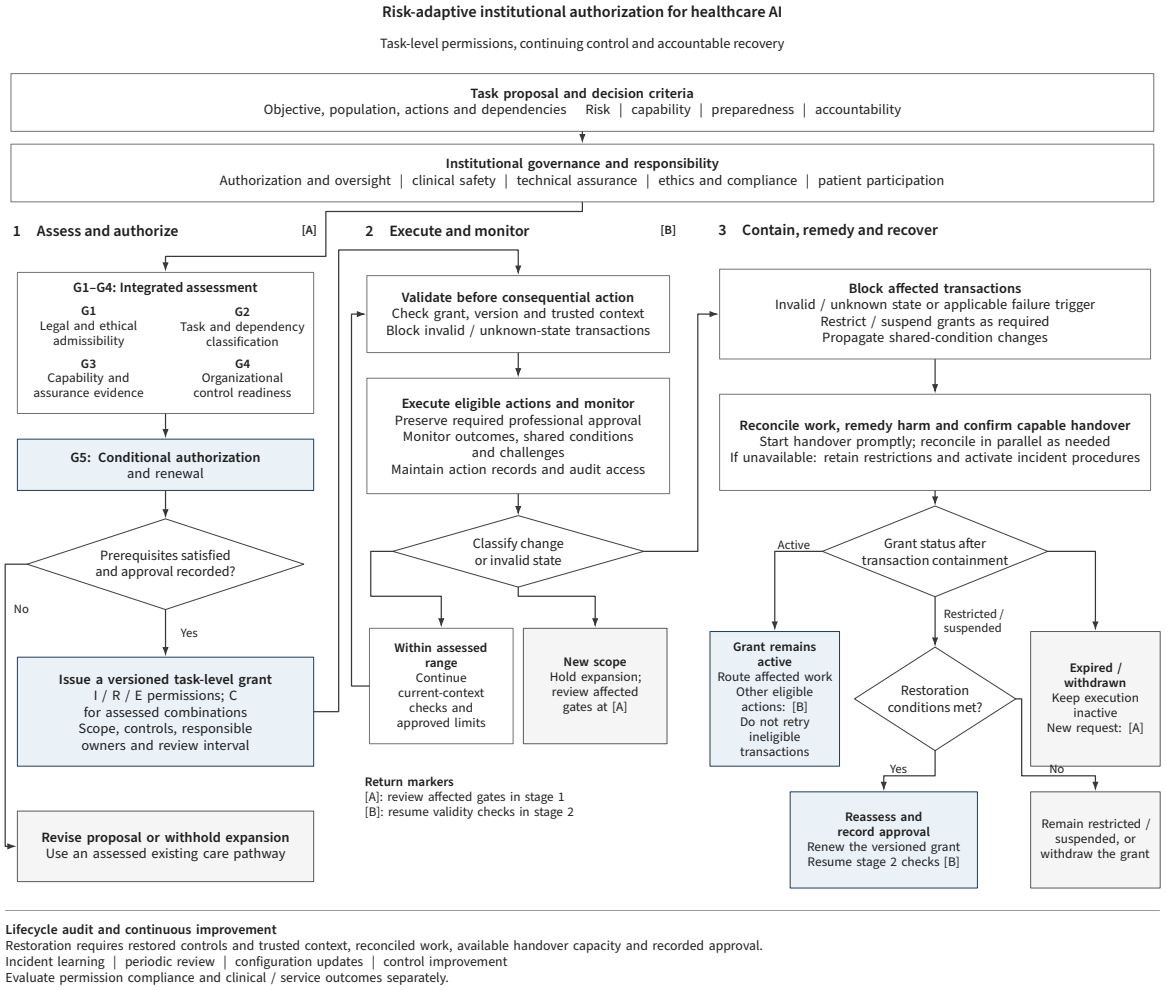


Figure 1. Task-level authorization and continuing control in RAGF

Note. Task proposals pass through five authorization gates informed by risk, capability, preparedness and accountability. Approved decisions yield versioned permissions for information support (I), recommendation (R) or bounded execution (E), with coordination (C) covering assessed combinations and permitted variations. Consequential execution requires a valid grant, configuration and trusted context. Within-range changes return to current-context checks; scope expansion awaits review of the affected gates. Invalid or unknown context blocks affected transactions, while applicable trigger rules determine grant-level restriction or suspension. Containment links work reconciliation and remedy to verified handover, which begins promptly while reconciliation proceeds in parallel as needed. If the grant remains active, affected work is routed and other eligible actions return to current-context checks; ineligible transactions are not retried. Restricted or suspended grants enter restoration review, while expired or withdrawn grants remain inactive and any new authorization request returns to the gates. Reinstatement requires restored controls and context, reconciled work, available handover capacity and recorded approval. Permission compliance and clinical or service outcomes are evaluated separately. Author's conceptual synthesis of Sections 4.2–4.4, 5.5 and 6.4.

4.5 Human oversight as a substantive capability

Effective human oversight requires task knowledge, interpretable evidence, sufficient time, and an intervention that reaches execution before credible harm occurs. Gate 4 assesses these capabilities under routine workload, interruption, and downtime. A recorded signoff establishes that a review step occurred; evaluation must also establish whether the reviewer could detect and prevent the relevant harm.

Prior approval controls consequential actions before execution; concurrent supervision supports intervention during delegation; retrospective audit detects trends and latent failures. Selection depends on consequences and recovery time. Clinical decisions can retain prior approval while bounded administrative actions use concurrent monitoring and audit.

Team evaluation tests error recognition, appropriate challenge, independent reasoning, and recovery during disconnection. Training uses counterexamples, conflicting information, and simulated dissent. Protected reporting allows staff to challenge advice and raise concerns without prior authorization.

4.6 Four theoretical propositions

The authorization rule sets out normative conditions for permission. The four propositions predict how particular governance arrangements affect allocation, workflow performance, reliance, and containment. Scenario analysis derives their decision implications; future empirical studies can assess the predicted mechanisms and outcomes. Legal obligations and patient rights provide the mandatory boundaries within which those arrangements are compared.

Proposition 1 (risk-sensitive allocation). Structuring the same case facts as an explicit map of irreversibility and downstream dependencies will reduce execution scope more for high-consequence tasks than for comparable low-consequence tasks. The proposed mechanism is increased salience of the consequences. Committee vignettes would hold factual information, capability evidence, and controls constant while changing presentation. Permission sets and prior-approval requirements measure allocation; an independent assessment of task adequacy and consequences is needed to judge whether the allocation is better.

Proposition 2 (preparedness as a conditioning factor). Above a common validity threshold, higher task performance will have a stronger association with workflow benefit where institutional preparedness is greater. Preparedness is measured before deployment through staffing coverage, configured interfaces, and fallback resources demonstrated in separate exercises. Task performance must retain comparable variation above the threshold. Deployment-period handover failure and safe-escalation time are proposed mediators; a prespecified workflow outcome, such as timely task completion, is the endpoint. Comparable tasks across sites can examine this interaction, with case mix, resources, and selective adoption treated as competing explanations.

Proposition 3 (accountability and calibrated reliance). An arrangement combining named decision ownership, inspectable action records, and protected challenge rights will improve discrimination between appropriate acceptance and rejection of AI advice relative to liability assignment alone. This predicts the package's total effect. In a team simulation, clinical materials, AI advice, decision time, and available stop tools remain comparable; record access and protected challenge procedures are

intervention components. Understanding of responsibility and use of the challenge route provide measures of the proposed mechanism. Independently adjudicated cases define appropriate reliance; component-specific effects require separate comparisons.

Proposition 4 (precommitted permission transitions). With the same signals, stop tools, empowered staff, workload and fallback resources, precommitted restriction or suspension rules will reduce unsafe execution after an actionable signal relative to discretionary permission changes. In both conditions, the same authorized staff initiate the transition through the same interface; they differ in whether they apply a preapproved rule or decide after the signal. The proposed mechanism is a shorter signal-to-decision interval and, where decision-to-enforcement latency is unchanged, a shorter total signal-to-enforcement interval. The simulation records detection, decision, and enforcement times separately, together with adherence to the assigned decision procedure. Task tracking continues through handover over a common exposure window, including false stops and interruption harm.

Table 3 distinguishes evidence about mechanisms from evidence of net benefit. P1 concerns allocation, P2 the conversion of task performance into workflow benefit, P3 team reliance, and P4 containment. Service adequacy and safety require outcome assessment alongside measures of execution scope and stopping speed. Common outcome windows and denominators relate unsafe actions to attempted tasks, accepted unsafe advice to unsafe advice presented, and interruption harm to affected work. Independent adjudication, workload and access measures, and adjustment for clustering support comparisons at committee, team, or institutional level.

Table 3. Operationalization and disconfirming observations for future tests

Proposition and contrast	Operational measures	Evaluation design	Observation challenging the proposition
P1: Same facts with or without explicit consequence mapping	Execution actions granted; prior-approval requirements; difference by harm and dependency condition.	Committee vignettes fixing facts, capability, and controls; blinded permission coding.	No risk-by-presentation interaction. Independent adequacy assessment judges allocation quality.
P2: Preparedness conditions capability-to-benefit conversion	Predeployment staffing, interfaces, and exercised fallback; task performance above threshold; handover mediators; timely completion.	Comparable tasks across sites or staged adoption; examine performance-by-preparedness association and selection.	No predicted interaction challenges P2; burden exceeding benefit challenges net value separately.
P3: Accountability package versus liability assignment	Appropriate acceptance/rejection against adjudicated cases; responsibility comprehension; challenge use.	Team simulation fixing advice, case evidence, time, and stop tools; record access and challenge procedure vary.	No better reliance discrimination challenges total effect; mechanism inference requires corresponding process evidence.
P4: Precommitted versus discretionary permission changes	Detection, decision, enforcement times; unsafe actions per attempted task; false stops and interruption harm.	Same signals, tools, human initiator, interface, workload and fallback; preapproved rule versus post-signal discretion. Common exposure and follow-up.	No shorter containment path challenges mechanism; offsetting interruption harm challenges net benefit.

Note. The measures and study designs are proposed for future evaluation. Clinical outcome adjudication and safety criteria require independent specification.

Evaluation protocols would prespecify primary outcomes, observation windows, meaningful decision margins, signal criteria, and subgroup denominators. Register audits can use an independently established dependency list to examine missed affected grants, propagation time and unassigned pending work after a shared-condition change. Comparisons of linked and separately maintained grants should also record mistaken restrictions and maintenance burden; controlled simulations can then examine behavioral consequences. This connects the model's integration claim to observable maintenance failures as well as individual permission decisions.

5. Structured Scenario Analysis

5.1 Scenario A: Administrative coordination and scheduling

Consider an agent that reconciles appointment slots, staffing, equipment capacity, transport needs, and patient preferences. The requested grant permits rescheduling of designated nonurgent appointments inside an approved time window. Cross-department coordination could reduce conflicting bookings and unused capacity. Clinical services retain control over priority classifications and time-sensitive pathways, while patients receive an accessible route to question a change.

The main pathway to harm is delayed or inequitable access. Repeated changes can disrupt a sequence of tests, shift burdens toward patients with limited digital access, or defer a clinically important follow-up. Risk classification therefore considers the appointment's clinical dependencies and the cumulative effects of rescheduling. A booking can be changed back administratively, while a lost opportunity for timely care may be difficult to recover.

An E grant would cover allowlisted transactions for designated nonurgent pathways, constrain repeated changes, and protect clinical priority fields. C would additionally cover the approved sequence linking appointments and transport. The operations lead is responsible for execution; the clinical pathway owner maintains eligibility. Evidence includes local simulation, subgroup access checks, recovery testing, and shadow comparison with schedulers using the same constraints. Moving a single appointment within the approved clinical window and existing appointment–transport relationship uses the existing E and C grants, subject to current eligibility. A proposal that adds an unassessed cross-service action or changes the approved dependency relationship requires review before that sequence executes.

Suppose operations proposes automatic changes across several visits to improve capacity use. The clinical owner rejects this scope because cumulative movement can disrupt a time-sensitive pathway, while a patient objects to repeated travel changes. Technical evidence and access controls satisfy their gates, yet the requested grant remains disputed. Under Section 6.1, the clinical boundary excludes the affected pathway; an independent forum examines the remaining alternatives and patient burden. Following that review, the authorizer may approve single-visit changes within a clinically justified window with patient review, or retain manual scheduling if the narrower grant fails the service objective. The decision record states why the selected scope is sufficient.

Execution creates a second decision point. The clinical service updates an urgency flag after an appointment enters the queue. A context check at commit rejects the outdated change and pauses its dependent transport request. Unstarted work is cancelled; an already sent message requires reconciliation; a completed cancellation requires prompt clinical assessment and restoration of

necessary access where feasible. The patient's nondigital challenge reaches the pathway service, which can correct the record, secure a suitable appointment, and notify the authorizer. Recurrent disruption reopens the grant. Lost clinical time remains an adverse outcome even if the corrected workflow fully respects its revised boundaries.

5.2 Scenario B: Multidisciplinary diagnostic support

Consider an agent that assembles longitudinal records, laboratory values, imaging reports, and specialist notes to propose a differential diagnosis and further investigations. The requested permission is to present a structured recommendation, identify supporting sources and unresolved discrepancies, and prepare draft requests. A qualified clinician authorizes the diagnosis and consequential investigations or treatment under the existing clinical process.

The study by Zhang et al. (2026) provides a relevant technical comparator: an on-premise agent used behavioral consistency and other reliability measures to support selective autonomy in retrospective diagnostic benchmarks. The evidence concerns routing and diagnostic performance in the evaluated configurations. An institutional grant additionally requires local population and workflow evidence, an explicit boundary between a draft and a submitted order, and a usable escalation process.

An R grant would specify clinical scope, provenance, review responsibilities, and deferral. Recommendations expose material missing information and source disagreement. Staff independently check consequential claims and distinguish generated explanations from verified facts. The governance body assesses task evidence; the clinical service monitors the human–AI pathway, including the delay and workload introduced by mandatory review.

The disturbance removes a decision-relevant input, yet the agent still produces a plausible, consistent answer. The expected response is a visible uncertainty flag and clinical review, with submitted orders requiring separate professional authorization. Retrieved instructions to export records should be blocked by access controls. Evaluation distinguishes a boundary breach from harm that occurs within an observed boundary: a clinician may properly review an R output and still accept a wrong diagnosis. Unsafe advice accepted, missed diagnoses, unnecessary investigations, delay, workload, and subgroup deferral therefore remain outcome measures even when every access check passes.

5.3 Scenario C: Emergency treatment coordination

Consider an agent supporting emergency coordination by assembling critical information, locating resources, and contacting the responsible team. Both erroneous action and avoidable delay can cause irreversible harm. The requested permissions therefore distinguish rapid communication from treatment authority and assess each against its own time window and fallback requirements.

The proposed grant combines I and R with E permission to transmit a clinician-activated alert, subject to limits on recipients, message fields, and escalation conditions. C covers the approved sequence of notification and acknowledgement. An agent-generated urgency assessment can request review; the clinical activation supplies the execution trigger. Prescribing, medication administration, invasive treatment, and clinical-order changes remain with the responsible professionals. Authorization must account for the consequences of false, duplicated, or late alerts and specify preventive controls where later correction would be ineffective.

In the first branch, a recipient is unavailable and acknowledgement is delayed. The service's predefined pathway specifies when to wait and when to escalate, balancing the risk of delayed response against duplicate activation and competing demands. The emergency clinical lead approves that sequence after pathway-specific assessment and simulation. The model can assemble and route information within the grant, while qualified responders determine care. Evaluation follows both acknowledgement and arrival of a capable responder.

In a second branch, failure occurs after one notification has been sent. A shared identity or messaging service then fails, invalidating both agent transmission and the nominal digital fallback. Further agent transactions are blocked; unsent and unacknowledged items remain identifiable through the locally available work record or the service's downtime procedure. The duty clinician assumes coordination through a separately maintained route, such as direct contact or an onsite emergency call, only if that route is actually available. Receipt, staffing, and ability to act must be confirmed. If these dependencies also fail, Gate 4 is unsatisfied and agent coordination remains suspended; the clinical incident lead manages the service emergency under existing emergency procedures. The framework thus identifies conditions under which the requested coordination grant is infeasible.

Restoration combines infrastructure recovery with clinical reconciliation. The clinical lead reconciles sent and pending alerts with the receiving teams, checks duplicate activation and missed response, and verifies a usable fallback. The authorizer then records renewed permission for the restored configuration and its dependency structure. A responder arriving too late is an adverse clinical outcome even if every permission check was obeyed. This branch distinguishes a successful stop from a viable transfer of responsibility and a successful care outcome.

5.4 Cross-scenario implications

The cases distinguish constraints on a single action from permission to compose actions. Scheduling makes cumulative burden and professional jurisdiction decisive; diagnosis separates compliant access from sound judgment; emergency coordination exposes common dependencies in containment and recovery. Table 4 records both boundary failures and adverse outcomes occurring within the boundary. The same model can consequently receive different permissions as workflow, population, dependencies, and intervention time change.

Table 4. Scenario stress tests and expected authorization responses

Test dimension	Administrative scheduling	Diagnostic support	Emergency coordination
Harm pathway	Delayed or inequitable access through repeated or mistimed changes.	Unsafe judgment and downstream investigations or treatment.	Injury through delayed communication or erroneous action.
Proposed grant	E for eligible bookings; C for approved appointment–transport dependencies.	R recommendations and drafts; professional authorization of orders.	I and R; E for clinician-activated alerts; C for their approved escalation.
Evidence and owner	Workflow, access, and recovery evidence; operations lead and clinical pathway owner.	Local clinical and human–AI evidence; responsible clinical service.	Time-critical escalation and continuity tests; emergency clinical lead.
Perturbation	Repeated-change dispute; urgency changes before queued action commits.	A key input is missing despite a consistent answer.	Notification partly sent; shared messaging and nominal digital fallback fail.
Expected response	Adjudicate scope; block stale action and dependent transport; resolve patient challenge.	Expose uncertainty, retain clinician review, and block order submission under the R grant; submission requires separate professional authorization.	Stop affected transactions; reconcile alerts; confirm capable transfer or find grant infeasible.
Suspension and handover	Assign pending work; correct records and restore necessary access; reconsider recurrent harms.	Restrict affected recommendations or tools and notify the clinical owner.	Use an actually available independent route; reconcile messages and verify capacity before restoration.
Boundary failure and adverse outcome	Breach: stale action executes. Within-boundary harm: necessary care delayed or repeated travel burden.	Breach: unauthorized order or export. Within-boundary harm: reviewed but wrong advice accepted.	Breach: invalid transmission continues. Within-boundary harm: capable response arrives too late.

Note. These are constructed analytical cases. Expected decisions follow from the proposed rules; outcome criteria can identify harm despite rule compliance.

5.5 Cross-cutting perturbations and containment

In each case, a configuration change, missing audit record, failed policy check, or infrastructure outage can disconnect a grant from its evidence. Monitoring detects outcome problems; version and context checks identify changed conditions; access controls stop affected actions. The register must propagate a shared-condition change to dependent grants and responsible services. Recovery assessment records where this chain breaks, including a correct alert followed by ineffective action.

The integrity of permissions and the adequacy of the service are separate evaluation criteria. Containment can work even when delayed care, diagnostic error, or inequitable access persists. False alarms and patient review requests therefore enter both grant reassessment and outcome evaluation. Revocation requires stopping further exposure and assigning the unfinished work and consequences to capable services.

6. Institutional Implementation Pathways

6.1 Establish an AI authorization and safety function

Hospitals can place authorization within an existing multidisciplinary AI or patient-safety function. A proposed division assigns the operations lead responsibility for task objectives and alternatives, the clinical pathway owner for clinical eligibility and boundaries, and safety, privacy, and technical roles for their respective prerequisites. An institutionally empowered authorizer selects among eligible arrangements and records the reasons. Professional and legal boundaries constrain that choice. Other hospitals can map these functions to their own governance structure.

Disagreement over a prerequisite keeps the requested expansion inactive while a competent, non-conflicted reviewer resolves the issue; care continues through an assessed existing pathway. Among eligible options, an independent safety or ethics forum reviews disputes about necessity, burden, or fairness, with patient input and recusal of interested decision makers. Its findings can require narrower scope, additional conditions, or reconsideration of approval. Clinical decisions remain with appropriately authorized professionals. The register links each decision and review to the affected grants and their owners.

Joint Commission–CHAI guidance (2025) provides an organizational foundation, while FAIR-AI specifies business ownership, risk-based review and an inventory-supported monitoring plan (Wells et al., 2025). Freeman et al. (2026) show how structured assessment and oversight can be embedded in existing healthcare governance and adjusted to organizational maturity. RAGF applies these institutional foundations to continuing task and composition grants. Product regulatory status concerns applicable market or use requirements; procurement concerns acquisition; local authorization concerns a specified workflow. The responsible professional retains authority for the individual care decisions specified in that workflow.

6.2 Authorize specified tools and actions

Independent enforcement requires control over policy and trusted inputs as well as separation of software. Clinical services maintain authoritative eligibility; authorized administrators maintain grants. The receiving tool checks the applicable grant state and version at execution, implementing

complete mediation (Saltzer & Schroeder, 1975). Shared identity, network, record, or vendor services remain common failure dependencies even when policy code is separate. The grant records these dependencies and whether alternative paths are actually available.

Reversibility concerns effects already produced. Unstarted work can be cancelled; partially completed sequences require reconciliation before any retry; completed disclosures or lost clinical opportunities require consequence assessment and remediation. Each grant identifies who handles work in each state. In the emergency failure branch, sent messages require recipient reconciliation, unsent work needs explicit reassignment, and missing acknowledgements remain unresolved until a capable service confirms responsibility.

Vendor agreements should secure audit access, incident notification, data protections, update controls, and interruption support. Material configuration or workflow changes outside the assessed range, or changes meeting a recorded reassessment trigger, reopen the affected gates. Within-range replanning follows the approved limits, review requirements and current-context checks; configuration-version mismatches continue to block affected execution. A change with implications for product status or legal duties also requires the applicable external process before the hospital considers a renewed grant.

6.3 Establish evidence thresholds and staged deployment

A staged pathway uses laboratory testing, retrospective assessment, simulation, shadow operation, and controlled prospective evaluation as task risk requires. You et al. (2025) organize implementation around safety, efficacy, comparative effectiveness and monitoring; these stages provide a basis for matching evaluation to deployment decisions. Each stage specifies evidence thresholds, stopping rules, and permissions. Packages differ by task but address its identified harms and recovery conditions.

Evaluation covers omissions, subgroup performance, deferral, human understanding, and downstream actions. Disturbances include conflicting orders, missing inputs, retrieved malicious instructions, outages, and correlated failures. Monitoring relates outcomes to case mix, configuration, staff behavior, and workload; independent event review distinguishes model, control, and workflow failures.

6.4 Design meaningful accountability and incident response

Action records preserve inputs, configuration, provenance, transactions, approvals, overrides, and timing under privacy and retention controls. Applying Bovens's (2007) actor–forum account, RAGF assigns the operational owner to explain execution, the authorizing body to explain the grant, and an independent forum to examine adequacy and the need for remedy. Patient and professional challenges can therefore question both an individual action and the authorization arrangement that enabled it.

Restriction blocks further affected transactions, identifies unfinished and completed actions, assigns capable receivers, preserves evidence, and reopens the relevant gates. Handover requires both acknowledgement and confirmation that the receiving service has the capacity to act. Failed receipt or capacity triggers the recorded escalation. Reinstatement requires evidence of restored policy enforcement, trusted context, reconciled pending work, and available handover capacity, followed by recorded approval. Renewal therefore depends on restored operational control as well as adequate model performance. Outcome monitoring covers the interruption period and any remedial care.

6.5 Address equity and organizational capacity

Shared validation, regional assurance, and procurement support can reduce duplicated effort. Local populations, staffing, infrastructure, and escalation capacity determine feasible grants. Resource-constrained hospitals can assess narrower scopes against justified service objectives while retaining

mandatory control conditions. This makes institutional capacity an explicit condition of authorization and directs attention to differences in the services that patients can access across hospitals.

Patient review must produce an actionable decision. Accessible digital and nondigital routes reach a named service able to assess urgency, pause a disputed change where appropriate, arrange professional review, correct records, and restore necessary appointments or other services. The service explains the outcome and any further route of appeal; recurrent errors or barriers prompt review of the affected grant. An incident-remediation role addresses harm already incurred. Task-specific disclosure, legally required consent, and feasible alternatives are determined at Gate 1. Emergency care proceeds under the applicable urgent-care pathway, followed by review.

7. Discussion: Theoretical Contribution and Future Research

RAGF links component permissions, system configuration, task dependencies, and cross-service ownership through continuing institutional authorization. Consider two arrangements with identical risk information, staff, stop tools, and component permissions. In one, each service maintains its grant separately; in the other, a shared eligibility change identifies and reopens every dependent permission and assigns outstanding work. The second arrangement changes which actions remain eligible and who must respond, even while individual tools continue to function. The distinction identifies an observable governance mechanism: shared-condition review propagates a change across permissions and assigns responsibility for its consequences. The propositions and proposed dependency-failure comparisons address its behavioral effects, maintenance burden, and net benefits.

Cross-domain capability makes this relationship more important by making unfamiliar task combinations easier to generate. The scheduling case shows why evidence for summarization, booking, and transport does not settle permission for their new sequence. Component adequacy and whole-workflow admissibility are distinct judgments. RAGF makes the latter explicit and maintains it through changes, disputes, and withdrawal. Comparisons with independently maintained controls can assess whether this linkage improves coordination and recovery at an acceptable cost.

The framework connects healthcare AI governance to public administration through the distribution of discretion. A grant determines whose decisions an agent can implement, whose work it can reorganize, and whose concerns can interrupt execution. Changes in model capability alter the evidence available to institutional decision makers. Professional mandates, patients' rights, and feasible controls determine the authority that those decision makers can confer.

Greater capability can support bounded scheduling while diagnostic decisions retain professional approval. Coordination also creates shared failure risks and can increase organizational dependence. Manual recovery, staffed alternatives, and independent communication impose continuing costs. Their demonstrated availability determines which authority is feasible; an unavailable alternative is a reason to restrict or withhold the requested grant.

Governance arrangements can themselves produce harm through delay, alert burden, unnecessary suspension, or concentrated decision power. RAGF therefore evaluates safety together with useful task completion, professional workload, patient access, and recovery costs. A mandatory review that routinely receives ritual approval offers little control; a sensitive trigger that repeatedly disrupts essential services may also worsen outcomes. The evaluation must identify an allocation of control that achieves the task objective within the specified safeguards.

The constructed cases establish analytical implications for authorization and recovery. Comparative and prospective studies are needed to estimate effectiveness and implementation costs. Local thresholds depend on clinical pathways, populations, applicable law, and available evidence. Rare catastrophic events, correlated vendor failures, and selective adoption complicate inference from routine experience. Evaluation should therefore combine ordinary workflow outcomes with disturbances to shared conditions and examine adverse outcomes that occur despite compliance with permission rules.

A practical research program begins with grant and dependency audits, followed by committee vignettes and team or workflow simulations. P2 can guide multisite or staged comparisons; P3 can first assess the accountability package and later separate components; P4 compares preapproved rules with post-signal discretion while holding the human initiator, interface and enforcement resources constant. Prospective studies require independent event adjudication and attention to case mix and adoption. Absent behavioral effects would challenge the specified mechanisms; interruption costs exceeding safety gains would challenge the net value of the tested arrangement. These comparisons assess design choices within applicable legal and professional constraints.

8. Conclusion

Healthcare institutions can prepare for more general AI by making authority over actions and their combinations explicit, reviewable, and revocable. RAGF connects four determinants to five gates and records permission for information, recommendation, bounded execution, and approved coordination. Cross-domain planning makes changing dependencies a central object of review. Professional boundaries, patient challenges, and independent adjudication shape which technically feasible tasks receive institutional authority.

The governing rule grants the least authority sufficient for a justified service objective within mandatory constraints. Versioned records and trusted execution checks connect that decision to action; shared-condition review, capable handover, and conditional restoration maintain it through change. The propositions specify how allocation, reliance, the conversion of performance into workflow benefit, and containment can be evaluated alongside burden and access. Institutional preparedness for AGI therefore includes the capacity to justify, enforce, contest, revise, and withdraw permission while sustaining care and addressing its consequences for patients and staff.

Declarations

Author Contributions

Feifei Chen: Conceptualization; Investigation; Writing—original draft; Writing—review and editing. **Chengwen Song:** Conceptualization; Methodology; Formal analysis; Visualization; Writing—original draft; Writing—review and editing. **Yuexin Lu:** Investigation (literature search). **Qilong Wang:** Supervision; Project administration; Writing—review and editing.

Funding

The authors received no funding for this study.

Conflict of Interest

The authors declare no conflicts of interest.

Acknowledgments

None.

Ethics Approval

This conceptual study analyzes published literature and constructed scenarios and involves no newly collected human-participant data or identifiable patient information. Ethics committee approval is not applicable to this research design.

Informed Consent

Informed consent is not applicable because this study involved no newly recruited human participants or identifiable patient data.

Data Availability

This conceptual study draws on published sources identified in the reference list. No new empirical datasets were generated or analyzed.

References

1. Bovens, M. (2007). Analysing and assessing accountability: A conceptual framework. *European Law Journal, 13*(4), 447–468. https://doi.org/10.1111/j.1468-0386.2007.00378.x
2. European Parliament & Council of the European Union. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Official Journal of the European Union, L*, 2024/1689. https://eur-lex.europa.eu/eli/reg/2024/1689/oj
3. Freeman, S., Wang, A., Saraf, S., Potts, E., Mckimm, A., Coiera, E., & Magrabi, F. (2026). From principles to practice: An actionable framework for AI governance in healthcare organisations. *BMJ Digital Health & AI, 2*(1), Article e000018. https://doi.org/10.1136/bmjdh-2026-000018
4. Goddard, K., Roudsari, A., & Wyatt, J. C. (2012). Automation bias: A systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association, 19*(1), 121–127. https://doi.org/10.1136/amiajnl-2011-000089
5. Goh, E., Gallo, R., Hom, J., Strong, E., Weng, Y., Kerman, H., Cool, J. A., Kanjee, Z., Parsons, A. S., Ahuja, N., Horvitz, E., Yang, D., Milstein, A., Olson, A. P. J., Rodman, A., & Chen, J. H. (2024). Large language model influence on diagnostic reasoning: A randomized clinical trial. *JAMA Network Open,* 7(10), Article e2440969. https://doi.org/10.1001/jamanetworkopen.2024.40969
6. Hager, P., Jungmann, F., Holland, R., Bhagat, K., Hubrecht, I., Knauer, M., Vielhauer, J., Makowski, M., Braren, R., Kaissis, G., & Rueckert, D. (2024). Evaluation and mitigation of the limitations of large language models in clinical decision-making. *Nature Medicine, 30*(9), 2613–2622. https://doi.org/10.1038/s41591-024-03097-1
7. Joint Commission & Coalition for Health AI. (2025). *Guidance on the responsible use of AI in healthcare (RUAIH)*. https://digitalassets.jointcommission.org/api/public/content/dcfcf4f1a0cc45cdb526b3cb034c68c2
8. Lekadir, K., Frangi, A. F., Porras, A. R., Glocker, B., Cintas, C., Langlotz, C. P., Weicken, E., Asselbergs, F. W., Prior, F., Collins, G. S., Kaissis, G., Tsakou, G., Buvat, I., Kalpathy-Cramer, J., Mongan, J., Schnabel, J. A., Kushibar, K., Riklund, K., Marias, K., . . . Starmans, M. P. A. (2025). FUTURE-AI: International consensus guideline for trustworthy and deployable artificial intelligence in healthcare. *BMJ, 388*, Article e081554. https://doi.org/10.1136/bmj-2024-081554
9. Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans, 30*(3), 286–297. https://doi.org/10.1109/3468.844354
10. Saltzer, J. H., & Schroeder, M. D. (1975). The protection of information in computer systems. *Proceedings of the IEEE, 63*(9), 1278–1308. https://doi.org/10.1109/PROC.1975.9939
11. Sittig, D. F., & Singh, H. (2010). A new sociotechnical model for studying health information technology in complex adaptive healthcare systems. *Quality and Safety in Health Care, 19*(Suppl. 3), i68–i74. https://doi.org/10.1136/qshc.2010.042085
12. Song, C., Hu, Y., Chen, F., Lenz, S., Li, J., Feng, Y., Zeng, X., & Wu, M. (2026). When algorithms become sacred: A cognitive-psychological model of trust, authority, and meaning in AI interaction. *Journal of Social Cognition and Communication, 1*(1), 22–31. https://doi.org/10.63802/jscc.V1.I1.291

13. Tabassi, E. (2023). *Artificial intelligence risk management framework (AI RMF 1.0)* (NIST AI 100-1). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.100-1

14. Vasey, B., Nagendran, M., Campbell, B., Clifton, D. A., Collins, G. S., Denaxas, S., Denniston, A. K., Faes, L., Geerts, B., Ibrahim, M., Liu, X., Mateen, B. A., Mathur, P., McCradden, M. D., Morgan, L., Ordish, J., Rogers, C., Saria, S., Ting, D. S. W., . . . the DECIDE-AI expert group. (2022). Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *Nature Medicine, 28*(5), 924–933. https://doi.org/10.1038/s41591-022-01772-9

15. Wang, A., Freeman, S., & Magrabi, F. (2026). Governance for safe and responsible AI in healthcare organisations: A scoping review of frameworks. *npj Digital Medicine, 9*(1), Article 516. https://doi.org/10.1038/s41746-026-02679-2

16. Wang, Q., Ma, S., Sun, J., Qi, X., Zhao, J., Ma, D., Liu, X., An, Q., Zhao, D., & He, S. (2026). Information quality, readability, and empathy of AI-generated public mental health information: A comparative evaluation of eight large language models. *Frontiers in Public Health, 14*, Article 1859078. https://doi.org/10.3389/fpubh.2026.1859078

17. Wells, B. J., Nguyen, H. M., McWilliams, A., Pallini, M., Bovi, A., Kuzma, A., Kramer, J., Chou, S.-H., Hetherington, T., Corn, P., Taylor, Y. J., Cuison, A., Gagen, M., Isreal, M., & FAIR-AI Consortium. (2025). A practical framework for appropriate implementation and review of artificial intelligence (FAIR-AI) in healthcare. *npj Digital Medicine, 8*(1), Article 514. https://doi.org/10.1038/s41746-025-01900-y

18. World Health Organization. (2021). *Ethics and governance of artificial intelligence for health: WHO guidance*. https://www.who.int/publications/i/item/9789240029200

19. World Health Organization. (2024). *Ethics and governance of artificial intelligence for health: Guidance on large multi-modal models*. https://www.who.int/publications/i/item/9789240084759

20. You, J. G., Hernandez-Boussard, T., Pfeffer, M. A., Landman, A., & Mishuris, R. G. (2025). Clinical trials informed framework for real world clinical implementation and deployment of artificial intelligence applications. *npj Digital Medicine, 8*(1), Article 107. https://doi.org/10.1038/s41746-025-01506-4

21. Zhang, L., Wölflein, G., Ferber, D., Liang, J., Carrero, Z. I., Wu, X., Vibert, J., Clusmann, J., Möhrmann, L., Möhrmann, E. E., Wichmann, C., Wolf, F., Lenz, T., & Kather, J. N. (2026). On-premise medical AI agents for reliable clinical decision-making. *Nature Medicine, 32*(10), 3705–3716. https://doi.org/10.1038/s41591-026-04609-x