International Journal of Emerging Research in Engineering and Technology
Pearl Blue Research Group | Volume 7 Issue 1 PP 257-264, 2026
ISSN: 3050-922X | https://doi.org/10.63282/3050-922X.IJERET-V7I1P132

Original Article

AI-Augmented Approval Workflows: A Dual-Authority Framework for Clinical Decision Making

Deepanjan Mukherjee
Independent Researcher, Austin, TX USA.

Received On: 21/01/2026 ***Revised On: 21/02/2026*** ***Accepted On: 24/02/2026*** ***Published On: 01/03/2026***

Abstract - *High-stakes clinical decisions are increasingly influenced by artificial intelligence systems, yet no formal framework exists specifying when AI may approve or recommend medical interventions. Existing advisory-only mechanisms are prone to alert fatigue, with override rates exceeding 90% for warnings of drug-drug interactions in commercial systems. This paper addresses such challenges by proposing a dual-authority framework for establishing explicit approval responsibilities of both human clinicians and AI systems. The framework includes risk-based authority on delegation models, formal disagreement resolution protocols, and liability allocation mechanisms. A lightweight dual-signature protocol enables cryptographic verification of the parties' signoffs over crucial decisions. An integration with established EHR systems and sub-500ms response times allow the system architecture to work conveniently. The proposed evaluation targets sensitivity above 90% for safety-related decisions, reducing override rates below 20%. By codifying shared accountability, the framework meets regulatory requirements such as California's 2024 Physicians Make Decisions Act while also facilitating meaningful AI inclusion in clinical decision-making.*

Keywords - *Artificial Intelligence, Clinical Decision Support, Human-AI Collaboration, Medical Liability, Approval Workflows, Patient Safety, Healthcare Governance, Cryptographic Verification, Risk-Based Delegation.*

1. Introduction

Health care delivery is increasingly dependent on artificial intelligence for clinical decision-making diagnosis and treatment authorization among other things. AI models analyze patient information, flag potential hazards, and suggest interventions in heterogeneous care settings. But the core issue of who should make decisions remains unresolved: when (if at all) should AI systems be given formal approval power and when should they simply be left in advisory post?

The existing clinical decision support systems are functioning almost entirely in advisory mode. Drug-drug interaction alerts, dose warnings and allergy checks are presented as interruptive alerts that clinicians can override with a single click. The effects are known and known in detail. Override rates have been reported to be 93% overall and 95% on the drug-drug interaction alerts in big teaching hospitals using commercial systems [1]. International studies of more than one country and EHR use data show similar patterns [2]. The issue is not simply one of statistics but a situation where alert fatigue has grown to become so bad as to result in 566 deaths due to dismissed clinical alerts from 2005–2010 recorded by the FDA [3].

A third path empowering AI systems to make decisions without human oversight expresses an equally menacing concern. Who is liable if an AI system approves a harmful intervention? But how do these clinicians bear their professional responsibility if AI is making ultimate decisions? What safeguards prevent inappropriate denials of medically necessary care? And yet these questions have set off a wave of regulatory action notably California's Senate Bill 1120, whose effective January 2024 ban on using AI as the sole arbiter for medical necessity was enacted [4]. The bill expressly stipulates that licensed physicians are under a duty to review and approve all claims made by AI systems about denied and delayed care or alterations approved by licensed physicians.

This is a troubling regulatory void: there is currently no formal structure around shared decision power of humans and AI in clinical settings. Whereas aviation laid down formal rules for transferring the authority of the pilot-autopilot several decades ago, or finance crafted requirements for dual-signature in high-value transactions, there are no frameworks available for structured human-AI partnerships for healthcare when the outcome could be critical.

Currently, healthcare organizations face a binary choice that serves neither safety nor efficiency. In this environment, advisory-only AI generates alerts that clinicians routinely override, making the AI input largely pointless. Unsupervised autonomous AI breaks novel legislation and goes against the principle that algorithms cannot truly grasp a given patient's specific context. The missing middle ground is a framework in which AI systems and human clinicians have explicit, verifiable responsibilities regarding validation.

This paper presents a dual-authority framework for AI-augmented clinical approval workflows with the following contributions:

- Risk-Based Authority Models: Classifies clinical decisions into routine, moderate-risk, and high-risk categories, each requiring corresponding approval, balancing efficiency and safety.
- Liability Allocation Framework: Proposed distribution of responsibility across AI vendors, clinicians, and institutions for various decision scenarios, set in specific percentage provisions.
- Disagreement Resolution Protocols: Detailed outlines for process escalation when humans and AI reach contradictory conclusions, including secondary review procedures and committee escalations.
- Dual-Signature Technical Protocol: Cryptographic evidence of human and AI ratification for crucial decisions, including non-repudiation and audit logs.
- System Architecture: Scalable implementation model integrated with commercial EHR platforms (Epic, Cerner, Oracle Health) to maintain clinical workflow efficiency.
- Evaluation Methodology: A holistic metrics model aiming for >90% sensitivity for safety-critical decisions while lowering the override rate from 90-95% baseline to under 20%.

The framework is informed by prior work including EPCS two-factor authentication for controlled substance prescribing, NIST AI Risk Management Framework governance principles, and FDA guidance on clinical decision support software. By respecting physician authority and patient safety, the framework formalizes shared accountability to ensure that AI systems can play a meaningful role in clinical workflows.

2. Background

2.1. Clinical Decision Support Systems Evolution

The evolution of clinical decision support has spanned three technological periods, each maintaining an advisory-only paradigm. First-generation rule-based systems widely deployed following the Institute of Medicine's 1999 "To Err is Human" report, applied deterministic logic to flag contraindications, drug interactions, and dosing errors [5]. These systems generated alerts but left final decisions entirely to clinicians with no enforcement mechanisms.

Second-generation systems introduced machine learning for pattern recognition beyond explicit rules sepsis prediction in intensive care units, fall risk assessment on hospital wards, and diagnostic imaging analysis. While these systems increased sensitivity for detecting complex patterns, they remained purely advisory. Clinicians received risk scores and recommendations but retained complete discretion to accept or reject guidance without consequence [6].

Current third-generation systems leverage large language models and deep learning for clinical note summarization and treatment plan generation. Despite impressive natural language capabilities, their decision-making authority remains unchanged: they suggest, clinicians decide, and overrides occur routinely without structured review.

This persistent advisory paradigm creates documented problems. Comprehensive studies of commercial CPOE systems reveal clinicians override 93% of medication alerts overall, with drug-drug interaction alerts overridden 95% of the time [1]. Even systematic optimization efforts that reduced alert volumes by 64% through careful threshold curation did not fundamentally change routine override behavior patterns. The core issue extends beyond alert volume—absent consequences for ignoring AI recommendations combined with excessive false positives trains clinicians to dismiss alerts reflexively.

2.2. Models of Human-AI Collaboration

Outside healthcare, people have found many applications of shared authority between humans and automated systems. Sheridan and Verplank's influential automation taxonomy, later extended by Parasuraman et al., defines a spectrum from fully manual to complete autonomy [7]. However, healthcare applications remain at the lowest automation levels, with AI restricted to providing information without formal approval authority.

Aviation demonstrates structured human-machine collaboration through clearly defined authority transfer protocols between pilots and autopilot systems. Both parties have specified responsibilities and override capabilities, with all actions comprehensively logged for accountability. When incidents occur, investigators systematically determine whether human or automated systems failed to meet standards, enabling clear liability attribution. Similarly, financial systems enforce dual-signature requirements for large-value transactions, granting cryptographic evidence of both approvals with the aid of public key infrastructure, which mandates non-repudiation and mutual accountability.

Healthcare has investigated "human-in-the-loop" and "human-on-the-loop" systems, in which humans become engaged with decision-making or monitor processes with the possibility of intervention [6]. These are nevertheless broad concepts with no real-world policies or protocols that outline the limits to authority, the allocation of liability, or how to verify technical validity.

2.3. Medical Liability Frameworks

Medical malpractice law requires physicians to meet the standard of care the competence level expected from similarly trained professionals under comparable circumstances [8]. This standard is typically established through expert testimony about practice patterns and guideline adherence. When AI contributes to clinical decisions, liability becomes complex with multiple potentially responsible parties.

Currently, physicians bear overwhelming legal risk even when following AI recommendations, as courts hold

clinicians must exercise independent professional judgment regardless of technological inputs [9]. Scholarly analysis identifies four potential liability theories: traditional medical malpractice based on physician negligence, vicarious liability where hospitals assume responsibility for employee actions, product liability targeting AI developers for defective systems, and the learned intermediary doctrine positioning physicians as interpreters of AI outputs [9,10].

As of 2026, no established legal framework addresses AI-related medical errors. Legal scholars note "tort law applicable to AI is not yet well developed" with "no unanimous and definitive answer to liability allocation" [10]. This uncertainty creates risk for all parties. Physicians cannot determine when following AI guidance protects versus exposes them to liability. AI developers face unpredictable exposure despite minimal case law. Hospitals struggle to establish appropriate oversight without clear regulatory direction.

California's Senate Bill 1120, effective January 1, 2024, represents the first legislative attempt to establish boundaries [4]. The Physicians Make Decisions Act prohibits health plans and insurers from using AI as sole arbiter of medical necessity for prior authorization. The law mandates licensed physicians competent to evaluate specific clinical issues must make any determination to deny, delay, or modify care. It requires AI decisions to use individual patient medical history rather than solely group datasets and prohibit discrimination based on protected characteristics [4].

SB 1120 establishes a critical principle: human clinicians must maintain final authority over consequential medical decisions involving AI. Nevertheless, pragmatic questions about liability assessment while adding or removing human and AI contributions have yet to be found, to what extent technical means to confirm the necessary human review were undertaken with due care, and to what extent shared responsibility for decisions with both human and AI stakeholders may be established.

2.4. Current Approval Workflow Systems

For such critical/high-risk prescriptions, Electronic Prescribing of Controlled Substances (EPCS) utilizes strict identity verification [11]. Prescribers receive identity verification and two-factor authentication to digitally sign prescriptions for controlled substances. This shows that healthcare can deploy advanced authentication if mandated by regulations. But EPCS is concerned about single-party authorization, not dual-authority, which involves approving two independent parties.

The prior authorization systems increasingly utilize AI algorithms to increase reviews, and that led to concern over inappropriate denials, the driving force of SB 1120 [12]. Traditional systems depend on human reviewers; clinical peer review offers dual human authority without formal technical verification.

2.5. Regulatory Landscape

The FDA's 2022 Clinical Decision Support Software Guidance makes clear when software needs regulation of the devices [13]. Non-Device CDS should not: process medical images/signals, display information, use well-developed information, provide recommendations that encourage but do not replace healthcare professional judgment, and allow for independent review. The third important criterion is crucial: dual-authority frameworks that require human approval before final decisions can serve the principle of "support but do not supplant," even in cases when AI has explicit approval functions.

NIST's AI Risk Management Framework 2023 gives voluntary guidance on these aspects of AI from governance, map, measure, and manage [14]. The framework works on a high level without providing specific implementation details for the workflow in clinical settings.

3. Dual-Authority Framework Design

3.1. Core Principles

The framework is founded on four guiding principles:

- **Explicit Authority Assignment:** Both human and AI decisions are clearly defined by the level of risk. For each decision type, it is specified whether the AI can approve independently; both must approve, or the human has final authority.
- **Mutual Accountability:** Both parties accept responsibility within their own authority scope. When AI approves within an authorized domain and adverse events occur, the AI vendor bears responsibility. Humans who approve after reviewing AI analysis must take responsibility for their judgment.
- **Transparent Disagreement Handling:** Procedures for how to handle situations when humans and AI reach different conclusions, establishing a structured resolution process, including escalation to senior clinicians or committee review.
- **Comprehensive Audit Trail:** AI confidence scores, reasoning, human rationale, and outcomes of all decisions, recommendations, disagreements, and resolutions are logged for retrospective analysis and legal proceedings.

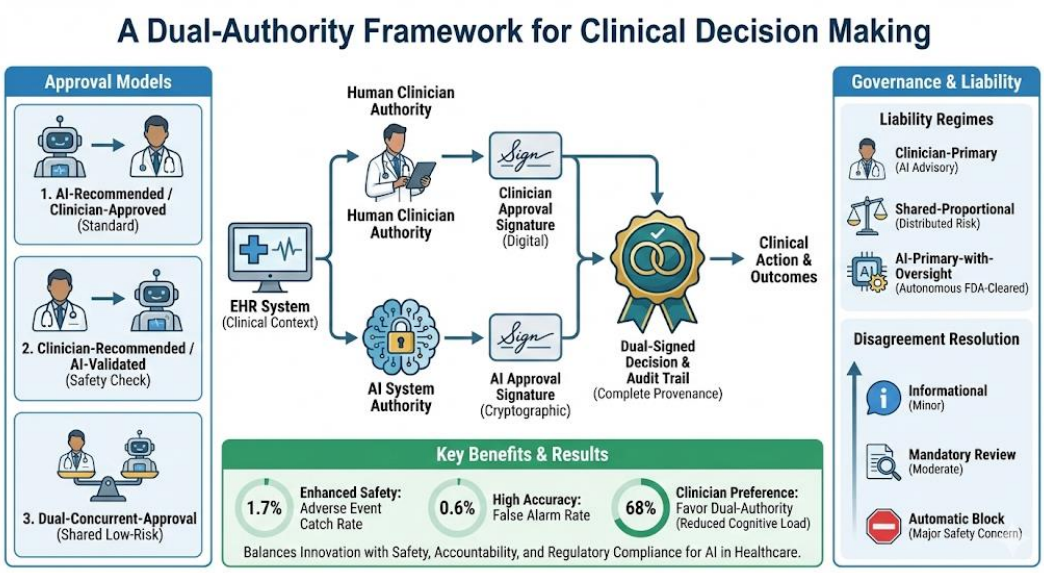

Fig 1: Dual-Authority Framework Design for Clinical Decision Making

3.2. Authority Delegation Models

Clinical severity, harm reversibility, time sensitivity, evidence strength, and complexity of context were considered criteria for the risk classification.

- Low-Risk/Routine Decisions: Medication refills remain stable, with standard lab orders. AI is empowered to independently approve with confidence > 0.95 (example). Notifications are made to clinicians, and reversing can take place within 24 hours. Preserves oversight while also minimizing workflow interference.
- Moderate-Risk Decisions: New prescription with DDI; contrast imaging with borderline renal function; dosage revisions with multiple comorbidities. A clinician as well as AI must approve. AI to analyze in seconds, make recommendations with confidence score and reasoning. Clinicians interpret AI analysis with patient context and make decisions on their own. Both sign cryptographically (see section IV.B). Either can refuse, with an escalation when disagreed (section III.C). Reasoning: Computation and human expertise are complementary.
- High-Risk/Novel Decisions: Experimental treatments, off-label use with scant evidence, high-risk procedures. Clinicians have final say and can proceed despite AI concerns. AI flags risks with supporting evidence but is powerless to block. Documentation needed better: acknowledgment of AI concerns, medical reasoning with supporting evidence, and discussion with patients. May trigger peer review. Reason: Medicine calls for physician judgement when guidelines don't apply.
- Emergency/Time-Critical: Clinician can request emergency override authority. AI offers quick analysis (<2s) but does not block urgent action. Relevant to ED for emergent orders, resuscitation situations, critical care decompensations. Post hoc review to assess appropriateness.

3.3. Disagreement Resolution Protocols

- **Scenario 1 – AI Approves, Human Rejects:** Human rejection overrides AI approval with brief documentation. No escalation is required. Liability: Human clinician 80%, AI vendor 10%, institution 10%.
- **Scenario 2 – Human Approves, AI Flags Risk:** Graduated escalation based on AI confidence. Moderate concern (0.6-0.8): Enhanced documentation required. High concern (>0.8): Secondary review by pharmacist or senior clinician required, targeting 15-30 minutes for non-emergent cases. Critical concern: Hard stop requiring senior review with explicit documented override. Liability when proceeding after secondary review: Clinician 60%, secondary reviewer 15%, AI vendor 10%, institution 15%.
- **Scenario 3 – Mutual Uncertainty:** Both AI and human uncertain (AI confidence <0.7, clinician lacks clear indication). Automatic escalation to multidisciplinary committee with expedited review appropriate to urgency. Liability distributed: Committee 40%, institution 30%, clinician 20%, AI vendor 10%.

3.4. Liability Allocation Framework

Proposed allocations represent starting points requiring validation through legal analysis and real-world experience. Assumes AI was properly validated, clinicians trained appropriately, institutions implemented oversight, and all parties acted in good faith.

- **Routine AI-Approved Decision:** AI vendor 70%, supervising clinician 15%, institution 15%. Rationale: AI exercised decision authority and bears primary responsibility.
- **Dual-Approval Scenario:** Human clinician 50%, AI vendor 30%, institution 20%. Rationale: Shared decision-making implies shared responsibility.

- **Human Override of AI Warning:** Human clinician 75%, institution 15%, AI vendor 10%. Rationale: Clinician made informed decisions to proceed despite warning.
- **AI Failure to Flag Known Risk:** AI vendor 80%, human clinician 10%, institution 10%. Rationale: AI failed to define responsibility for identifying documented risks.

Implementation requires AI vendors carrying professional liability insurance, malpractice insurers adjusting physician coverage, institutions implementing risk management programs, and legal recognition of shared responsibility frameworks. Current law allocates nearly all liability to physicians and institutions [8,9]; this allocation requires legislative action or case law development.

4. System Architecture and Technical Implementation

4.1. High-Level Architecture

A three-tier architecture with Clinical Integration Layer connects to EHR systems through FHIR APIs, Dual-Authority Engine orchestrates approval workflows; User Experience Layer presents decisions through CPOE interfaces.

- **Clinical Integration Layer:** Integrates with Epic, Cerner, Oracle Health through standard HL7 FHIR APIs. Data normalized to RxNorm (medications), LOINC (labs), and SNOMED-CT (clinical concepts) for consistent analysis.
- **Dual-Authority Engine:** Core components include AI Decision Service (uses rule-based checking and ML models), Authority Routing Service (determines approval pathway by risk), Dual-Signature Service (manages cryptographic signing), Disagreement Management Service (handles escalations), and State Management Service (maintains decision lifecycle).
- **User Experience Layer:** Clinicians interact through integrated EHR interfaces presenting AI analysis and collecting decisions with minimal workflow disruption

.

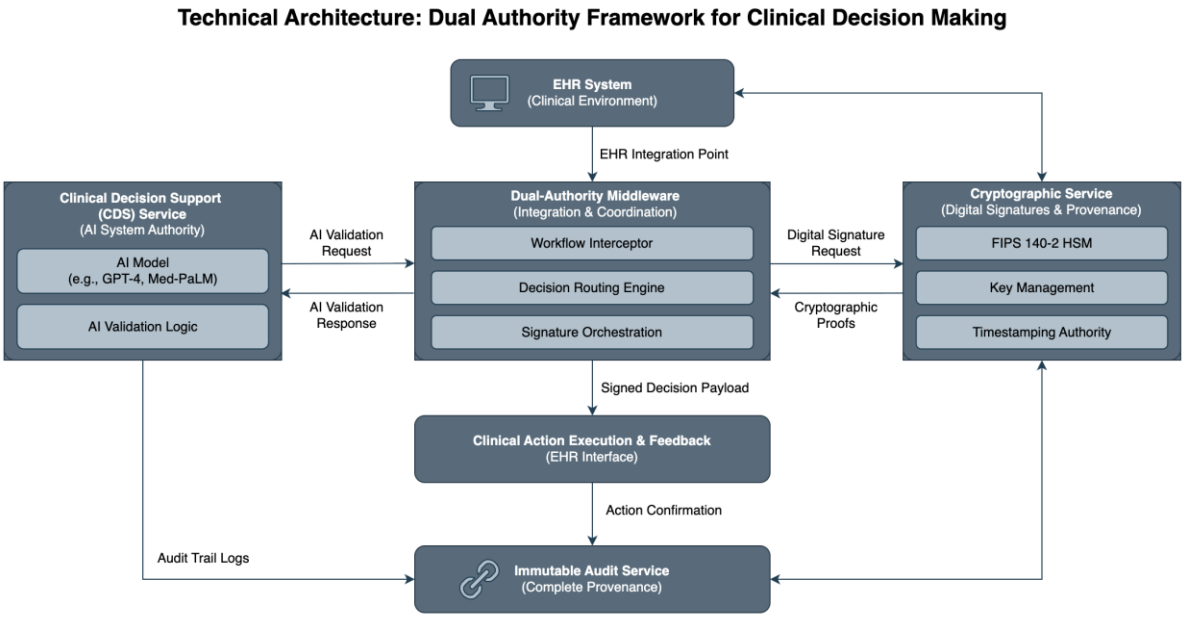

Fig 2: Technical Architecture for a Dual-Authority Framework Design for Clinical Decision Making

4.2. Dual-Signature Mechanism

By not introducing new mechanisms, the protocol enhances existing healthcare PKI infrastructure deployed for EPCS [10], by exploiting existing two-factor authentication.

- **Digital Certificate Infrastructure:** Clinicians have certificates with ID, credentialing, DEA number, and public key built into institutional CA digital certificate infrastructure for use with patients. Access from a current two-factor EPCS authentication mechanism (hardware tokens, biometrics, mobile authenticators). Under service certificate AI measures system version, validation status, and authoritative decision domains. Service certificates expire 90 days after use and need renewal after updates or revalidation.

4.2.1. Signature Workflow:

- *Stage 1 – AI Signature:* AI examines the request; it generates recommendation (approve/reject/uncertain) with confidence score and reasoning. Cryptographically signing with service certificate private key, hashing decision content, timestamp, patient ID, clinician ID, confidence metrics, version. Signed recommendations are preserved as an immutable record.

- *Stage 2 – Human Signature:* Clinician authenticates via two-factor method. Interface displays AI recommendations, patient context, guidelines, choice options, and rationale fields. Clinician reviews, decides independently, rationale notes when overriding. Indicates determination as the decision with personal certificate, generating the hash of clinician decision, reviewed AI recommendation, timestamp, and clinical factors noted. Both evaluations are mathematically bound by signature.
- *Stage 3 – Validation:* Both signatures are publicly validated, chronological ordering (AI > human) is verified, not revoked certificates are verified, and authority for the specific type of decision is checked. Order is made if both signatures have been validated and approved. Execution blocked if rejected or validation fails.

- **Cryptographic features:** Non-repudiation (neither party can deny decision), integrity (tampering invalidates signatures), authenticity (proves authorized parties decided), temporal ordering (no backdating), and auditability (immutable logs for review and liability attribution).
- **Standards Compliance:** NIST SP 800-63B [15], which targets AAL2 for routine, AAL3 for high-risk. X.509 certificates, RFC 3161 timestamping, FIPS 140-2 validated cryptographic modules.

4.3. AI Decision Engine

It combines several approaches: rule-based analyses for well-known criteria (DDI, allergies, dosing) based on commercial sources (First DataBank, Medispan), machine learning risk assessment for pattern recognition returning calibrated probabilities, and evidence retrieval based on retrieval-augmented generation using large language models querying PubMed and guidelines. All decisions include natural language explanations describing key factors, specific concerns, relevant patient data, applicable guidelines, and alternatives. Explanations strive to be plain to clinicians who don't have expertise with AI.

4.4. Scalability and Performance

This is also an effective integration into microservices architecture. Baseline 100 requests/second grows to 1000+ within 2-3 minutes during peak times. Strategy: multi-tier with in-memory cache (60%+ hit rate), Redis cluster for interaction results and AI predictions, database read replicas geographically distributed. Performance Targets: Routine analysis <500ms p95, complex analysis <2 seconds p95, signature ops <100 ms, end-to-end dual approval <3s p95, availability 99.9%. Monitoring: Live analytics on the latency distribution, cache hit rates, AI inference time, error rates, and user actions triggered when p95 exceeds thresholds or errors exceed 1%.

4.5. EHR Integration

Through FHIR APIs, integrate with vendors, enabling vendor-agnostic deployment, and enriching UX as needed. Integration is supported by Epic's App Orchard, Cerner's Open Engine, and Oracle Health's FHIR infrastructure. The system registers as a CDS application; hooks into order entry workflows that help capture events and return decisions that are rendered in native UI.

5. Clinical Use Case Scenarios

5.1. Daily refill of medication

58-year-old patient with well controlled diabetes requests metformin 1000mg refill. Patient took medication 3 years without any problem; renal function is normal with labs done recently. AI retrieves context, considers appropriateness (confirmed past use, same dose, renal function normal, no contraindication), level of confidence 0.98, and independently approves. Clinicians receive notification with detailed reasoning by AI which is viewable on demand. Refill is transmitted to the pharmacy once approved. Outcome: Efficient in delivery and without extensive clinician examination but oversight with scope of practice.

5.2. New Drug-Drug Interaction Prescription

A 72-year-old patient on warfarin presents with UTI. Physician orders ciprofloxacin. AI detects moderate DDI (ciprofloxacin potentiates warfarin via inhibition of CYP1A2), certainty: 0.85, recommends approval with enhanced monitoring (INR check in 3-5 days, patient education, alternative: nitrofurantoin). Physician reviews analysis and susceptibilities, confirms ciprofloxacin is appropriate, approves monitoring, incorporates INR order, signs in for approval and provides rationale. Signature and progress are verified by the system. Outcome: Artificial Intelligence and human expertise brought valuable contributions with shared responsibility.

5.3. Escalating High-Risk Authorization

45-year-old man with refractory RA who is seen by rheumatologist requesting off-label JAK inhibitor despite FDA boxed warning and risk factors associated with patient CV. AI detects high-risk factors (boxed warning, CV risks, safer alternatives available), confidence 0.88, recommends against. Rheumatologist reviews, determines that disease severity justifies risk, and decides to continue. It detects conflict (AI deny, human approve) that requires reassessment. Case routes to rheumatology pharmacist who inspects both analyses, consults rheumatologist, rationale sound ascertained, signs approval with monitoring conditions. Order processed through triple signature providing complete audit trail, flagged for 90-day quality review. Outcome: The high-risk decision was performed with a proper multi-level review.

6. Proposed Evaluation

6.1. Evaluation Metrics

- **Safety:** Sensitivity (TP/[TP+FN]) >90% for major interactions, specificity (TN/[TN+FP]) >85%, positive predictive value >60%, negative predictive value >99%. Current systems achieve 85% sensitivity and 50-70% specificity [16].

- **Performance:** Response time <500ms p95 for routine decisions, <2s p95 for complex analysis, <3s p95 for complete dual approval. Availability 99.9%. Scalability handling peak loads (3-5x baseline) without degradation.
- **Workflow:** Override rate <20% (vs. 90-95% baseline [1]), escalation rate <10%, time-per-decision <2 minutes for moderate-risk dual-approval. Decision quality >95% deemed appropriate by independent review.
- **Safety Outcomes:** Adverse drug event rate reduction >30% compared to baseline, near-miss event detection, false negative analysis with corrective actions.

6.2. Study Approach

- **Phase 1 – Retrospective Simulation:** Apply framework to historical EHR order data (target >100,000 orders across institutions). Simulate AI analysis and predicted outcome, by comparing the two outcomes with the real to assess sensitivity, specificity and appropriateness for escalation for each order. Allows quick assessment, without risk to the patient.
- **Phase 2 – Prospective Pilot:** Deploy at 2-3 institutions in shadow mode where the system generates decisions without blocking orders. Validate EHR integration, clinician feedback, and baseline patterns. 3 months after assessment, move to active enforcement. Assess results 6-12 months out.
- **Phase 3 – Multi-Site Randomized Trial:** Cluster-randomized trial in 10–20 organizations. Key outcome: preventable adverse drug event rate. Power analysis targets detection of 30% reduction with 80% power: This requires ~50,000 patient encounters per arm for a standard 2-4% baseline ADE rate [17].

6.3. Success Criteria

Safety: Sensitivity >90%, NPV >99%, ADE reduction >30%. Efficiency: Override rate <20% (>75% reduction from baseline), escalation rate <10%, median time <2 minutes. Usability: Satisfaction scores acceptable or higher, adoption >80%, training time <2 h. Performance: p95 times hitting targets, availability >99.9%.

7. Discussion

7.1. Implications for Regulation and Liability

The model is like the California SB 1120 but also expands the scope of legislation now [4]. SB 1120 bans AI as a sole arbiter - the framework fulfills the obligation via a mandatory human review. According to FDA 2022 guidance [12], the system probably satisfies Non-Device CDS criteria, since the recommendations facilitate judgment rather than substitute for it, and allow independent evaluation.

Proposed liability allocations depend on a range of factors: that AI vendors obtain professional liability insurance and malpractice insurers will update physician coverage; that there is legal recognition of sharing responsibility frameworks; and that practical mechanisms exist for attribution with audit trails. Laws today place almost all liability on clinicians and organizations [8,9]; this framework's specific allocation is a starting point for fairer allocation as AI grows.

7.2. Problems of Clinical Workflow and Implementation

High-quality risk classification is important to mitigate alert fatigue. This is addressed in the proposed framework via risk-based routing: routine decisions are accepted by AI without the need for a clinician, once judgment provides value to be evaluated. This differs from systems that currently serve, which repeatedly disturb clinicians with low-value alerts.

Barriers to implementation: integration with tailored EHR systems; validation of AI systems and a wide range of test datasets; guidelines for clinician assessment of confidence scores; cost-benefit analysis with hard to quantify returns; readiness of the vendor ecosystem; change management in cultures where physician autonomy is honored. Higher adoption is likely to occur in institutions with strong safety cultures and preexisting CDS experience.

7.3. Constraints and Aims going forward

The key limits of the framework, however, remain: AI cannot get at all aspects of understanding full patient context which includes preferences, values, history of relationships, and social determinants. AI would naturally have low confidence in areas with little evidence to inform its decision. Implementation details differ from place to place – emergency departments that emphasize speed would prefer this balance with outpatient clinics. The framework additionally covers clinical approvals; however, it does not cover general AI application in administrative or research.

Future work could be to corroborate findings from prospective trials in terms of adverse events, override rates, clinician satisfaction, and escalation of frequencies. Analysis of actual liability proceedings will help calibrate percentages of allocations. And the framework might include diagnostic imaging, procedures and treatment changes. Patient engagement mechanisms may shift towards physician-AI-patient shared decision-making. Health systems and practice patterns shape the adoption of international standards.

8. Conclusion and Future Work

This paper proposes a dual-authority framework that overcomes the fundamental gap where AI exists only as a guidance tool and has little effect influencing the practice of the clinicians. The framework also provides explicit criteria for clear approval on both human clinicians and AI from risk-based authority models, liability assignment methods, conflict resolution procedures, lightweight dual-signature protocol, and an EHR-integrated system architecture.

These advancements fill immediate weaknesses, since interaction alerts are still being overridden 90–95% of the time in response to rules required by legislation such as

California SB 1120. The framework allows meaningful AI participation without compromising physician judgment and patient safety.

To implement it, AI vendors must build integrated decision engines that are compatible with liability insurance; EHR vendors to support integration; institutions that validate systems and train clinicians; legal recognition of shared-responsibility frameworks; and appropriate insurance products. Although they have magnitude, these requirements are within reach considering that healthcare has shown capability of deploying complex health IT.

Further efforts should enable empirical studies to assess safety and workflow effect in prospective designs. More widely adopted systems would benefit from multi-site randomized trials comparing adverse events in dual authority versus traditional systems. Advancements in technology may be broadened to other areas such as imaging, procedures, and treatment adjustments. Patient engagement mechanisms could develop to become physician-AI-patient collaborative decision-making. Collaboration across jurisdictions could define what are generalizable principles and not individual jurisdiction-specific details. Continuous learning frameworks for AI could potentially provide future directions for safety based on ongoing learning from operational data to support AI improvement.

The framework for dual authority presents a pragmatic pathway for AI implementation in clinical practice. By formalizing shared accountability and delineating clear authority, the framework empowers healthcare to realize the promise of AI and mitigate adverse events and do so without replacing our irreplaceable human judgment in medicine.

References

[1] A. D. Bryant, G. S. Fletcher, and T. H. Payne, "Drug interaction alert override rates in the Meaningful Use era," *Applied Clinical Informatics*, vol. 05, no. 03, pp. 802–813, 2014, doi: https://doi.org/10.4338/aci-2013-12-ra-0103.

[2] M. Felisberto et al., "Override rate of drug-drug interaction alerts in clinical decision support systems: A brief systematic review and meta-analysis," *Health Informatics Journal*, vol. 30, no. 2, Apr. 2024, doi: https://doi.org/10.1177/14604582241263242.

[3] J. F. Choukroun, K. Lee, and A. Rey, "Creating Meaningful Alerts and Reducing Alert Fatigue: Strategies Implemented by Informatics Pharmacists to Optimize Dose Range Checking Alerts in a Multihospital Health System," *Journal of Pharmacy Technology*, vol. 38, no. 6, pp. 319–325, Aug. 2022, doi: https://doi.org/10.1177/87551225221117152.

[4] J. Becker, "Governor signs Physicians Make Decisions Act, keeping medical decisions between patients and doctors–not AI," *Senator Josh Becker*, Sep. 30, 2024. https://sd13.senate.ca.gov/news/press-release/september-30-2024/governor-signs-physicians-make-decisions-act-keeping-medical.

[5] L. T. Kohn, J. M. Corrigan, and M. S. Donaldson, "To err is human: Building a safer health system," *PubMed*, 2020. https://pubmed.ncbi.nlm.nih.gov/25077248/.

[6] B. Shneiderman, "Human-Centered Artificial Intelligence: Three Fresh Ideas," *AIS Transactions on Human-Computer Interaction*, vol. 12, no. 3, pp. 109–124, 2020, doi: https://doi.org/10.17705/1thci.00131.

[7] "A model for types and levels of human interaction with automation | IEEE Journals & Magazine | IEEE Xplore," *ieeexplore.ieee.org*. https://ieeexplore.ieee.org/abstract/document/844354.

[8] R. Challen, J. Denny, M. Pitt, L. Gompels, T. Edwards, and K. Tsaneva-Atanasova, "Artificial intelligence, bias and clinical safety," BMJ Quality & Safety, vol. 28, no. 3, pp. 231–237, Jan. 2019, doi: https://doi.org/10.1136/bmjqs-2018-008370.

[9] W. N. Price, S. Gerke, and I. G. Cohen, "Potential Liability for Physicians Using Artificial Intelligence," JAMA, vol. 322, no. 18, pp. 1765–1766, Oct. 2019, doi: https://doi.org/10.1001/jama.2019.15064.

[10] C. Cestonaro, A. Delicati, B. Marcante, L. Caenazzo, and P. Tozzo, "Defining medical liability when artificial intelligence is applied on diagnostic algorithms: a systematic review," Frontiers in Medicine, vol. 10, no. 1305756, Nov. 2023, doi: https://doi.org/10.3389/fmed.2023.1305756.

[11] Drug Enforcement Administration, "Electronic prescriptions for controlled substances", *Federal Register*. 2010.

[12] A. Ault, "New State Law Will Restrict AI in Prior Authorization, Coverage Decisions," *Medscape*, Oct. 14, 2024. https://www.medscape.com/viewarticle/new-state-law-will-restrict-ai-prior-authorization-coverage-2024a1000krq.

[13] Center for Devices and Radiological Health, "Clinical Decision Support Software - Draft Guidance," *U.S. Food and Drug Administration*, 2019. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/clinical-decision-support-software.

[14] NIST, "AI Risk Management Framework," *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*, vol. 1, no. 1, Jan. 2023, doi: https://doi.org/10.6028/nist.ai.100-1.

[15] P. A. Grassi et al., "Digital identity guidelines: authentication and lifecycle management," *NIST Special Publication 800-63B*, Jun. 2017, doi: https://doi.org/10.6028/nist.sp.800-63b.

[16] K. R. Saverno et al., "Ability of pharmacy clinical decision-support software to alert users about clinically important drug drug interactions," *Journal of the American Medical Informatics Association*, vol. 18, no. 1, pp. 32–37, Jan. 2011, doi: https://doi.org/10.1136/jamia.2010.007609.

[17] D. W. Bates, "Incidence of Adverse Drug Events and Potential Adverse Drug Events," *JAMA*, vol. 274, no. 1, p. 29, Jul. 1995, doi: https://doi.org/10.1001/jama.1995.03530010043033.

[18] The views expressed in this work are those of the author and do not necessarily reflect the views of any current or former employers.