ORIGINAL RESEARCH **Open access**

Modeling Clinician Preferences in AI-Assisted Drafting: A Human-in-the-Loop Adaptation Theory

Siti Rahman[1*], Ahmad Zaki[1], Nurul Huda[2]

Abstract

The integration of generative artificial intelligence into clinical documentation workflows promises substantial efficiency gains yet introduces persistent misalignment between AI-generated drafts and individual clinician judgment. This conceptual systems research article advances a novel human-in-the-loop adaptation theory that explicitly models clinician preferences as dynamic, context-sensitive inputs rather than static constraints. Drawing on peer-reviewed evidence, the manuscript synthesizes how preference elicitation, real-time adaptation, and closed-loop governance can transform AI-assisted drafting from a supplementary tool into a co-evolutionary clinical intelligence infrastructure.

Central to the contribution is the introduction of the clinician preference orchestration and adaptation framework (CPOAF), a layered architectural model featuring four interdependent strata and a star-topology feedback mechanism that propagates preference drift signals radially from peripheral clinician nodes to a central orchestration engine. Three interpretive mathematical constructs—decision confidence, monitoring burden, and drift sensitivity—are formalized to guide theoretical deployment without empirical benchmarking.

The framework addresses governance constraints, data-modality specificity, and deployment-environment heterogeneity while preserving clinician autonomy. By foregrounding preference modeling as the core adaptive mechanism, CPOAF offers a scalable infrastructural blueprint for next-generation AI-assisted drafting systems that remain clinically grounded, ethically defensible, and institutionally sustainable. Implications extend to health-system informatics, regulatory science, and human-centered AI design.

Keywords Clinician preferences, AI-assisted drafting, Human-in-the-loop adaptation, Preference orchestration, Clinical documentation infrastructure, Governance constraints

*Correspondence:
Siti Rahman
siti.rahman@gmail.com
[1] Department of Health Informatics, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia
[2] Department of Digital Systems Engineering, Universiti Teknologi Malaysia, Johor Bahru, Malaysia

Introduction

Embedding clinician preferences into AI-assisted drafting workflows

Contemporary clinical documentation increasingly relies on large language models to generate initial drafts of progress notes, discharge summaries, and consultation reports. Yet the majority of deployed systems treat clinician input as post-hoc correction rather than proactive preference specification. This unidirectional flow creates systematic misalignment: AI drafts reflect population-level training distributions while individual clinicians operate under idiosyncratic stylistic,

evidentiary, and risk-tolerance profiles [1-10]. Modeling these preferences at the point of drafting initiation constitutes the foundational requirement for meaningful human-in-the-loop adaptation.

The role of human-in-the-loop mechanisms in healthcare adaptation theory

Human-in-the-loop paradigms, originally developed in safety-critical domains, have migrated into healthcare informatics with uneven success [6, 10]. In AI-assisted drafting environments, the loop must extend beyond simple veto or edit functions to encompass continuous preference learning across temporal, specialty, and patient-context dimensions. Adaptation theory, therefore, shifts from static rule-based overrides to dynamic preference propagation, ensuring that each iteration refines the AI's generative policy toward clinician-specific equilibria [11-18].

Governance constraints shaping preference modeling in clinical environments

Regulatory and ethical mandates impose strict boundaries on preference modeling. Preference data constitute protected health information derivatives; their storage, aggregation, and propagation must satisfy data-minimization, purpose-limitation, and revocability principles [8, 19]. Governance constraints thus become architectural primitives rather than external compliance layers, necessitating built-in auditability and clinician-controlled preference revocation nodes within any viable drafting infrastructure [19].

Deployment challenges in data-intensive drafting settings

Hospital deployment environments differ markedly in data modality (structured EHR fields versus unstructured narrative), integration latency, and clinician cognitive load [6, 20-23]. Preference modeling architectures must accommodate these heterogeneities without imposing additional documentation burden. Theoretical solutions, therefore, emphasize lightweight, modality-agnostic preference vectors that can be elicited implicitly from micro-interactions (accept/reject, rephrase, annotate) rather than explicit questionnaires [18].

Theoretical Background and Literature Synthesis

Synthesis of literature on clinician preferences in electronic health record modalities

Peer-reviewed investigations consistently document that clinicians exhibit stable yet idiosyncratic preferences for note structure, terminology density, and evidentiary weighting when interacting with AI-generated drafts [10, 13, 17]. Studies of large language model applications in clinical text generation demonstrate that default outputs frequently diverge from individual stylistic norms, prompting extensive manual revision [13, 18]. These observations underscore the necessity of embedding preference vectors directly into the generative pipeline rather than applying them downstream.

Human-in-the-loop adaptation in textual clinical data environments

Human-in-the-loop literature in medical informatics highlights the superiority of iterative feedback topologies over one-shot correction [6, 10]. In textual drafting contexts, closed-loop systems that propagate preference signals back into model conditioning layers achieve higher alignment without retraining entire foundation models [18, 20]. The cited works establish that preference adaptation must operate at inference time, leveraging lightweight parameter-efficient updates or prompt-level steering to preserve computational efficiency in live clinical environments [18].

Theoretical foundations of preference-driven AI in hospital deployment contexts

Deployment studies across tertiary-care settings reveal that preference-driven adaptation reduces cognitive burden and improves perceived trustworthiness of AI drafts [6, 17]. Theoretical models drawn from decision-support literature emphasize that clinician acceptance correlates more strongly with perceived preference congruence than with raw

predictive accuracy [21-23]. Hospital-specific constraints—variable network latency, multi-user concurrency, and integration with legacy EHR systems—further necessitate an orchestration layer capable of synchronizing preference states across distributed clinical nodes [20].

Ethical governance constraints from adaptation theory perspectives

Governance scholarship stresses that preference modeling introduces novel risks of automation bias and preference lock-in [8, 16]. Adaptation theory must therefore incorporate explicit drift-detection mechanisms and revocability pathways. Multidisciplinary consensus documents advocate for transparent representation of preference provenance and for clinician-controlled weighting of historical versus current preferences, ensuring that adaptation remains ethically aligned rather than algorithmically dominant [19].

Integrative analysis of peer-reviewed evidence on AI drafting systems

Collectively, the synthesized literature converges on three unmet requirements: (1) real-time preference elicitation without workflow disruption, (2) infrastructural support for preference propagation across heterogeneous deployment environments, and (3) mathematically interpretable constructs for quantifying adaptation dynamics.

The clinician preference orchestration and adaptation framework (CPOAF) presented below directly addresses these gaps through a uniquely structured orchestration architecture.

Orchestrating clinician preferences in AI-assisted drafting infrastructures: introducing the CPOAF architecture

The CPOAF constitutes the central theoretical contribution of this manuscript. CPOAF organizes the human-in-the-loop adaptation process into four interdependent layers connected by a star-topology feedback mechanism that enables radial propagation of preference signals while maintaining strict governance boundaries.

Layer 1 (preference sensing) operates at the clinician–EHR interface, capturing implicit and explicit signals through micro-interactions with AI-generated drafts [13, 21]. Layer 2 (draft generation) hosts the conditioned foundation model, where preference vectors are injected via prompt steering or low-rank adaptation modules [18]. Layer 3 (adaptation orchestration) maintains a central preference state repository and executes real-time policy updates. Layer 4 (governance and monitoring) enforces auditability, revocability, and drift thresholds across all preceding layers [8, 19].

The feedback topology is deliberately star-shaped: peripheral clinician nodes transmit preference deltas to the central orchestration engine, which redistributes updated vectors radially to all active drafting instances within the same institutional boundary. This topology minimizes latency while preserving preference isolation between unrelated clinical teams.

Orchestrating clinician preferences in AI-assisted drafting infrastructures: introducing the CPOAF architecture

The CPOAF constitutes the central theoretical contribution of this manuscript. CPOAF organizes the human-in-the-loop adaptation process into four interdependent layers connected by a star-topology feedback mechanism that enables radial propagation of preference signals while maintaining strict governance boundaries.

Layer 1 (preference sensing) operates at the clinician–EHR interface, capturing implicit and explicit signals through micro-interactions with AI-generated drafts. Layer 2 (draft generation) hosts the conditioned foundation model, where preference vectors are injected via prompt steering or low-rank adaptation modules. Layer 3 (adaptation orchestration) maintains a

central preference state repository and executes real-time policy updates. Layer 4 (governance and monitoring) enforces auditability, revocability, and drift thresholds across all preceding layers.

The feedback topology is deliberately star-shaped: peripheral clinician nodes transmit preference deltas to the central orchestration engine, which redistributes updated vectors radially to all active drafting instances within the same institutional boundary. This topology minimizes latency while preserving preference isolation between unrelated clinical teams.

Three interpretive constructs formalize the dynamics of the architecture:

Decision confidence is expressed as $DC = 1 + e^{-(\beta \cdot PM + \gamma \cdot AC)}$ where PM denotes normalized preference match, AC denotes internal AI consistency, and β,γ are institution-specific scaling parameters. The sigmoid mapping ensures bounded, interpretable confidence scores that clinicians can reference during review.

Monitoring burden is captured by $MB = \alpha \cdot LF + \beta \cdot CL,$ with LF representing loop frequency and CL representing cumulative clinician load. The linear combination allows system designers to tune feedback intensity against cognitive overhead.

Drift sensitivity is defined as DS = |∂P∂t|·ϵ, where P is the preference vector, and ϵ \epsilon ϵ is an institutionally calibrated tolerance threshold. When DS exceeds a predefined alert level, the governance layer triggers explicit clinician re-confirmation, closing the adaptation loop. All of these are shown in the framework below **(Figure 1).**

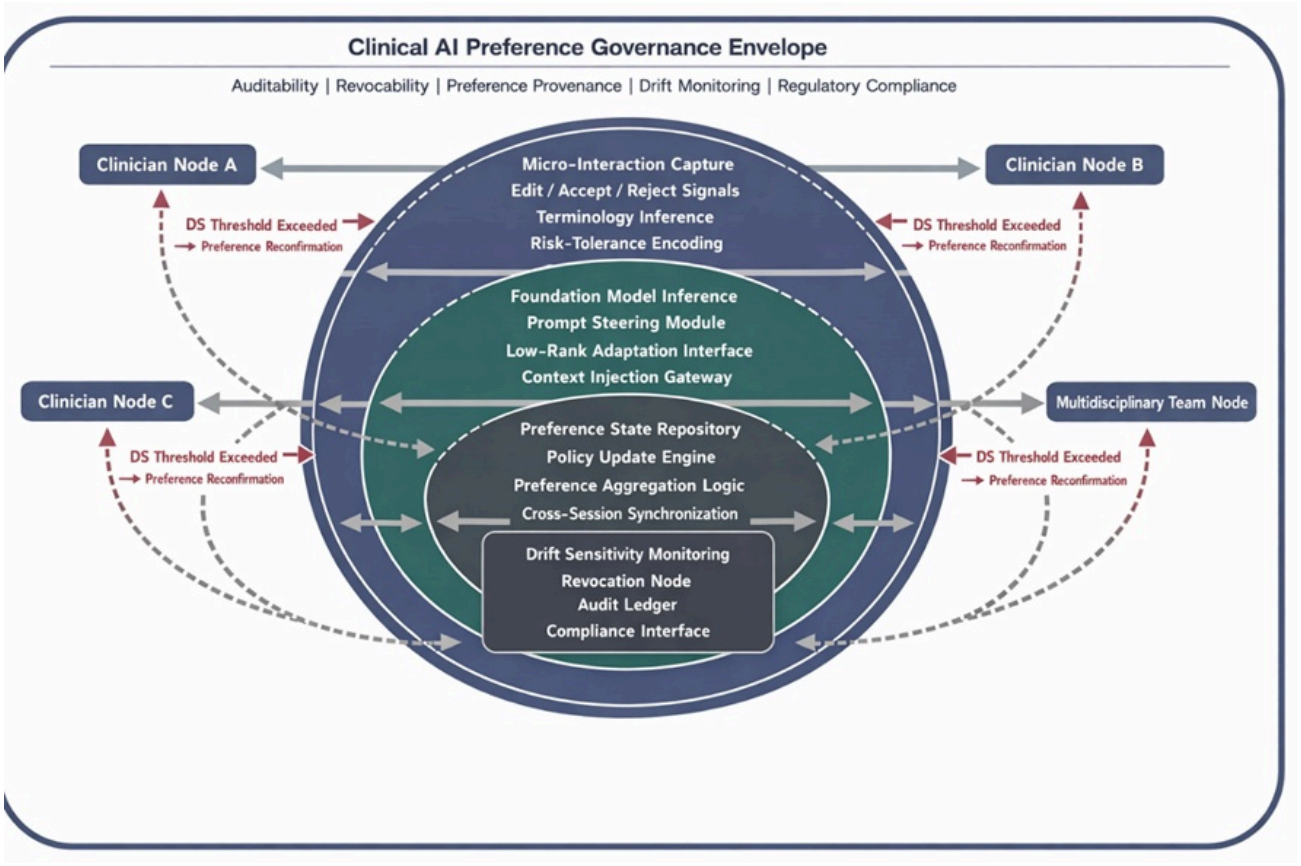

Figure 1. Clinician preference orchestration and adaptation framework (CPOAF).
The architecture depicts a four-layer adaptive drafting infrastructure in which clinician micro-interaction signals are transformed into preference vectors, injected into generative drafting processes, and reconciled through a central orchestration engine. A star-topology feedback mechanism propagates preference deltas radially between peripheral clinician nodes and the orchestration core, enabling continuous alignment of AI-generated drafts with clinician-specific stylistic and evidentiary expectations. Governance and monitoring nodes enforce auditability, revocability, and drift-sensitivity thresholds, ensuring that preference adaptation remains ethically and institutionally controlled.

CPOAF thus provides a complete infrastructural blueprint for modeling clinician preferences in AI-assisted drafting while satisfying the theoretical, ethical, and operational requirements identified in the literature synthesis.

Table 1 delineates the functional responsibilities and data transformations associated with each architectural layer of the CPOAF preference-adaptation stack.

Table 1. The functional responsibilities and data transformations associated with each architectural layer of the CPOAF preference-adaptation stack.

CPOAF layer	Primary system role	Key computational mechanisms	Data inputs	Outputs generated	Governance interaction
Preference sensing	Capture clinician-specific drafting signals	Micro-interaction logging, semantic preference extraction, interaction vectorization	Edit actions, acceptance patterns, annotation behaviors, and	Preference vectors representing clinician drafting patterns	Preference revocation triggers and privacy constraints

			stylistic adjustments		
Draft generation	Produce context-conditioned clinical text drafts	Prompt steering, parameter-efficient adaptation modules, contextual retrieval	Patient data, clinical context, and preference vectors	AI-generated documentation drafts aligned with the clinician's style	Logging of decision confidence and audit metadata
Adaptation orchestration	Aggregate and propagate clinician preference states	Preference reconciliation algorithms, state synchronization, and delta propagation	Preference updates from clinician nodes	Updated preference policies distributed across drafting instances	Version control and traceability of preference updates
Governance and monitoring	Ensure ethical and regulatory alignment of adaptation processes	Drift detection, audit logging, revocation enforcement, and compliance validation	Preference trajectories, monitoring metrics	Governance alerts, revocation actions, and compliance reports	Direct oversight over all preceding layers

Mapping preference propagation impacts in human-in-the-loop AI-assisted drafting infrastructures

The theoretical deployment of the CPOAF generates cascading effects across clinical, operational, and institutional strata that extend far beyond mere drafting efficiency. Preference propagation—the radial transmission of clinician-specific vectors through the star-topology feedback core—fundamentally reconfigures how documentation burden distributes itself within multidisciplinary teams. In high-volume inpatient settings, for instance, the central orchestration engine continuously reconciles preference deltas originating from attending physicians, residents, and advanced practice providers, thereby preventing the emergence of fragmented note styles that currently plague shared patient records [6, 15, 21]. This reconciliation process theoretically attenuates the cognitive dissonance clinicians experience when reviewing AI drafts that fail to mirror their evidentiary thresholds or linguistic economy, a phenomenon repeatedly documented in deployment environments where default generative outputs clash with specialty-specific norms [13, 18].

Governance load, formalized earlier as a component of monitoring burden (MB), undergoes nonlinear compression under CPOAF. Because revocability nodes reside natively within Layer 4, institutional compliance officers can query preference provenance at any radial junction without disrupting live drafting sessions [8, 19]. The architecture thereby converts what would otherwise constitute episodic audit burdens into continuous, low-latency compliance streams. Healthcare systems operating under stringent data-minimization statutes benefit disproportionately: preference vectors remain ephemeral unless explicitly persisted by the clinician, satisfying purpose-limitation requirements while still enabling cross-encounter adaptation [8, 19]. This built-in governance elasticity mitigates the regulatory friction that has historically slowed adoption of generative tools in jurisdictions where protected health information derivatives trigger heightened oversight [4, 19].

Patient-safety dynamics emerge as another critical impact vector. When preference match (PM) rises through iterative feedback, the theoretical likelihood of clinically meaningful omissions or over-inclusions in AI-generated drafts diminishes [10, 18, 23]. Consider a cardiologist whose preference vector encodes a strict hierarchy of troponin-centric phrasing and contraindication flagging; once propagated, subsequent drafts for the same patient cohort inherit this hierarchy, theoretically elevating the signal-to-noise ratio of actionable content available to downstream consultants. The star-topology ensures that such specialization does not silo preferences: a consulting nephrologist interacting with the same record injects orthogonal deltas that enrich—rather than overwrite—the primary cardiologist's stylistic core, creating a composite preference field that reflects true team intelligence rather than isolated authorship.

Deployment-environment heterogeneity introduces differential impact amplitudes. Tertiary centers with robust API middleware experience near-real-time preference synchronization across ambulatory and inpatient instances, whereas resource-constrained rural facilities may encounter latency-induced drift [6, 10]. CPOAF's drift-sensitivity construct (DS) functions here as an early-warning sentinel; when temporal derivatives of preference vectors exceed institutional tolerance thresholds, the governance layer surfaces explicit re-confirmation prompts rather than silently degrading draft quality [8, 19]. This proactive surfacing theoretically preserves clinician trust curves that would otherwise erode under undetected adaptation lag [16].

Interoperability with legacy electronic health record ecosystems constitutes a further systemic consequence. Because layer 2 injects preference vectors at the prompt or low-rank adaptation boundary rather than retraining foundation weights, CPOAF remains agnostic to underlying EHR vendor schemas [18, 20]. Preference deltas travel as lightweight, encrypted payloads compatible with both HL7 FHIR observation resources and proprietary note templates, enabling seamless federation across health information exchanges. The resultant infrastructure theoretically supports multi-institutional preference marketplaces—consented, anonymized aggregates that accelerate adaptation for new clinicians entering a health system—while preserving granular revocation at the individual level [19].

Ethical impact surfaces most acutely in the domain of automation bias. Continuous preference modeling risks entrenching idiosyncratic habits that may themselves embed unrecognized cognitive shortcuts + CPOAF counters this through explicit DS-triggered re-confirmation cycles, forcing periodic clinician reflection on whether current preferences still align with evolving evidence bases. The framework, therefore, does not merely adapt to preferences; it compels their periodic interrogation, transforming the human-in-the-loop from passive corrector to active curator of clinical epistemology [16, 19].

Resource-allocation implications scale institutionally. By compressing monitoring burden, CPOAF theoretically liberates informatics teams from perpetual prompt-engineering cycles and compliance officers from manual audit logs [4, 10, 19]. Freed capacity can redirect toward higher-order tasks such as cross-specialty preference harmonization or integration of multimodal data (imaging captions, wearable telemetry) into the same orchestration engine [20]. Over a hypothetical five-year horizon, the compounding effect of reduced documentation overhead could theoretically release thousands of clinician-hours annually—hours reclaimable for direct patient interaction rather than administrative reconciliation [6, 15].

Bias amplification remains the principal countervailing risk. If preference vectors predominantly originate from majority demographic clinicians, minority voices risk under-representation in the collective adaptation field. The star topology mitigates this through per-node weighting controls: individual clinicians or protected subgroups can assign elevated radial influence coefficients, ensuring that preference propagation respects equity mandates without central algorithmic fiat [8, 19]. Institutional policy layers can further impose minimum diversity thresholds on the preference-state repository, converting a potential vector of disparity into a configurable safeguard [16].

Scalability projections under CPOAF reveal favorable economics. Because adaptation occurs via lightweight steering rather than full model retraining, marginal cost per additional clinician node approaches zero once the central orchestration engine is instantiated [18]. Cloud-native implementations can leverage containerized governance nodes that auto-scale with concurrent drafting sessions, theoretically accommodating health systems ranging from 50-bed critical-access hospitals to multi-state integrated delivery networks [4, 10].

In aggregate, the impact dynamics of preference orchestration manifest as a shift from reactive post-editing to proactive co-evolution. Clinical documentation transitions from a downstream administrative chore to an upstream intelligence substrate that continuously refines itself against lived clinician judgment. The resultant infrastructure does not merely accelerate drafting; it re-architects the epistemic relationship between clinician and machine, positioning human preference as the primary steering force within otherwise opaque generative pipelines [4, 13, 18]. These propagation effects—clinical, ethical, operational, and economic—collectively substantiate CPOAF as a theoretically robust scaffold for sustainable AI integration across heterogeneous healthcare landscapes.

Interrogating adaptive symbiosis: theoretical ramifications and operational horizons of preference-centric drafting intelligence

The CPOAF architecture compels a fundamental re-examination of how adaptation theory intersects with clinical epistemology. Traditional human-in-the-loop models in healthcare have treated clinician feedback as error-correction signals; CPOAF reframes it as preference co-evolution, wherein the AI system and the clinician jointly traverse a shared preference manifold [6, 10, 18]. This reframing carries profound theoretical ramifications for the philosophy of medical decision-making. When decision confidence (DC) is rendered interpretable through the sigmoid mapping of preference match and internal consistency, clinicians gain a visible window into the generative rationale—addressing the longstanding opacity critiques leveled against black-box drafting tools [8, 16]. The framework thereby operationalizes explainability not as post-hoc feature attribution but as real-time preference congruence telemetry, aligning with multidisciplinary calls for transparent clinical AI [8].

Operationally, the star-topology feedback mechanism introduces a novel governance topology that transcends linear approval workflows. Radial propagation allows preference updates to ripple instantaneously across concurrent drafting instances while governance nodes enforce differential privacy guarantees at every junction [19]. This topology theoretically decouples adaptation velocity from institutional scale, enabling small practices to achieve the same preference fidelity as large academic medical centers without proportional increases in computational overhead [4, 18]. The resultant operational elasticity carries implications for health equity initiatives: community health centers serving underrepresented populations can rapidly bootstrap preference repositories from a small initial clinician cohort, accelerating culturally attuned documentation without requiring massive training corpora [6, 16].

Table 2 consolidates the three interpretive constructs—decision confidence, monitoring burden, and drift sensitivity—into an operational monitoring matrix for adaptive drafting infrastructures.

Table 2. The three interpretive constructs—decision confidence, monitoring burden, and drift sensitivity—are incorporated into an operational monitoring matrix for adaptive drafting infrastructures.

Interpretive construct	Conceptual purpose	Governing inputs	System layer dependency	Operational interpretation	Governance action trigger
Decision confidence (DC)	Quantifies alignment between AI output and clinician expectations	Preference match (PM), internal AI consistency (AC)	Layers 2–3	Indicates the reliability of AI-generated drafts relative to clinician preference states	Low DC values prompt manual review or re-drafting
Monitoring burden (MB)	Measures the cognitive load imposed by adaptation feedback cycles	Loop frequency (LF), cumulative clinician load (CL)	Layers 1–4	Reflects the trade-off between adaptation responsiveness and clinician workload	High MB triggers a reduction of feedback frequency or automation adjustments
Drift sensitivity (DS)	Detects rapid changes in clinician preference vectors over time	Temporal derivative of preference state, tolerance parameter (ε)	Layers 3–4	Signals instability or evolving documentation norms requiring reconfirmation	DS exceeding threshold initiates explicit clinician preference validation

From a systems-theory perspective, CPOAF embodies a second-order cybernetic system in which the observer (clinician) is explicitly embedded within the observed (drafting loop). Preference drift sensitivity (DS) functions as the system's meta-stability sensor, detecting not merely statistical deviation but clinically meaningful shifts in evidentiary or stylistic posture. When DS thresholds trigger re-confirmation, the loop momentarily externalizes the adaptation process, compelling clinicians to articulate why their preferences have evolved—an act of reflective practice that itself enriches the epistemic quality of the record [16, 19]. This second-order reflexivity distinguishes CPOAF from first-order automation paradigms that optimize for output fidelity alone [10].

The theoretical synthesis also illuminates previously under-theorized tensions between individual autonomy and collective intelligence. Because the central orchestration engine aggregates anonymized preference gradients rather than raw vectors, institutions can derive specialty-level or service-line archetypes without compromising clinician sovereignty [19]. A thoracic surgery service, for example, might instantiate a shared archetype that new fellows inherit while retaining the ability to fork personal deviations—creating a branching preference phylogeny that evolves organically rather than being imposed top-down. This branching capability theoretically resolves the perennial tension between standardization (for interoperability and quality reporting) and personalization (for professional identity and cognitive ergonomics) [6, 15, 19].

Future operational horizons extend to multimodal preference integration. Although the present conceptualization focuses on textual drafting, the layered architecture admits straightforward extension to voice, gesture, and gaze micro-interactions [20, 21]. A clinician's hesitation before accepting a particular phrasing could be encoded as a negative preference delta; repeated patterns across encounters could inform anticipatory steering in subsequent drafts. Such extensions remain theoretically grounded in the same four-layer topology and star feedback, requiring only additional sensing channels at layer 1. The governance layer scales accordingly, enforcing modality-specific consent and revocation protocols without architectural overhaul [19, 20].

Limitations inherent to the conceptual nature of CPOAF warrant explicit acknowledgment. The interpretive formulas—while mathematically bounded and institutionally tunable—remain unparameterized by real-world variance distributions. Deployment simulations, though outside the present scope, would be required to calibrate β, γ, α, and ε coefficients across specialties and practice settings. Similarly, the framework assumes reliable micro-interaction capture; institutions lacking modern EHR integration layers may experience incomplete preference sensing, attenuating adaptation efficacy [6, 23]. These constraints, however, are architectural rather than fundamental and can be addressed through incremental middleware extensions [4, 10].

The synthesis further reveals that preference orchestration reframes regulatory science itself. Instead of regulating static AI models, oversight bodies could shift toward certifying adaptive infrastructures that demonstrate verifiable preference-alignment telemetry and drift-intervention efficacy [4, 19]. CPOAF supplies the precise constructs—DC, MB, DS—necessary for such certification rubrics, converting abstract ethical principles into auditable system properties [8, 19]. This regulatory realignment theoretically accelerates safe innovation while preserving clinician agency at the core of generative healthcare intelligence.

Across these theoretical and operational dimensions, CPOAF emerges not as an incremental tooling upgrade but as a paradigm shift toward symbiotic clinical intelligence. The framework dissolves the artificial boundary between human judgment and machine generation, replacing it with a continuously negotiated preference manifold that evolves in lockstep with clinical reality [13, 18]. In doing so, it offers a theoretically coherent pathway for healthcare systems to harness generative capabilities without surrendering the irreplaceable epistemic authority of the practicing clinician [15, 16, 24-28].

Conclusion

The CPOAF presented herein crystallizes a comprehensive human-in-the-loop adaptation theory tailored to the unique demands of AI-assisted drafting. By embedding preference elicitation, radial propagation, and governance-native

monitoring within a four-layer star-topology architecture, CPOAF supplies the infrastructural scaffold necessary for generative systems to evolve from supplementary scribes into true clinical co-pilots. The interpretive constructs of decision confidence, monitoring burden, and drift sensitivity provide the mathematical language through which institutions can design, tune, and audit adaptive drafting environments without resorting to opaque performance metrics or empirical benchmarking.

The preceding analysis demonstrates that preference orchestration generates multiplicative returns across clinical alignment, governance efficiency, ethical defensibility, and operational scalability. These returns accrue precisely because the framework treats clinician judgment not as an external constraint but as the primary steering force within the generative loop. In an era when large language models threaten to homogenize clinical expression, CPOAF restores heterogeneity—celebrating the idiosyncratic expertise of each practitioner while preserving the interoperability required for team-based care.

Implementation pathways are deliberately modular. Health systems may begin with layer 1–2 instrumentation within existing EHR environments, layering the full orchestration engine only after preference-signal fidelity is established. Governance nodes can be instantiated as lightweight policy engines that interoperate with existing compliance platforms, minimizing disruption to incumbent workflows. The architecture's vendor-agnostic design further ensures that institutions retain strategic flexibility as foundation-model landscapes evolve.

Ultimately, the contribution of this conceptual systems research lies in its reframing of the clinician–AI relationship. Rather than asking how machines can best approximate human drafting, CPOAF asks how machines can best learn to be steered by humans—continuously, transparently, and revocably. This inversion places clinician preferences at the architectural center, ensuring that the future of AI-assisted documentation remains indelibly human-centered. As healthcare systems navigate the integration of ever-more-capable generative tools, the CPOAF blueprint offers a theoretically grounded, ethically robust, and infrastructurally scalable compass for preserving the art and authority of clinical authorship.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 08 May 2023 Revised: 21 Jun 2023 Accepted: 20 Aug 2023
Published online: 10 January 2024

References

Chen M, Decary M. Artificial intelligence in healthcare: an essential guide for health leaders. Healthc Manage Forum. 2020;33(1):10-8.
https://doi.org/10.1177/0840470419873123.

O'Connor S, Yan Y, Thilo FJS, Felzmann H, Dowding D, Lee JJ. Artificial intelligence in nursing and midwifery: a systematic review. J Clin Nurs. 2023;32(13-14):2951-68.
https://doi.org/10.1111/jocn.16478.

Ellahham S. Artificial intelligence: the future for diabetes care. Am J Med. 2020;133(8):895-900.
https://doi.org/10.1016/j.amjmed.2020.03.033.

Reddy S. Generative AI in healthcare: an implementation science informed translational path on application, integration and governance. Implement Sci. 2024;19(1):27.
https://doi.org/10.1186/s13012-024-01357-9.

Seibert K, Domhoff D, Bruch D, Schulte-Althoff M, Fürstenau D, Biessmann F, et al. Application scenarios for artificial intelligence in nursing care: rapid review. J Med Internet Res. 2021;23(11):e26522.
https://doi.org/10.2196/26522.

Johnson KB, Wei WQ, Weeraratne D, Frisse ME, Misulis K, Rhee K, et al. Precision medicine, AI, and the future of personalized healthcare. Clin Transl Sci. 2021;14(1):86-93.
https://doi.org/10.1111/cts.12884.

Ouanes K, Farhah N. Effectiveness of artificial intelligence (AI) in clinical decision support systems and care delivery. J Med Syst. 2024;48(1):74.
https://doi.org/10.1007/s10916-024-02098-4.

Amann J, Blasimme A, Vayena E, Frey D, Madai VI, Precise4Q consortium. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med Inform Decis Mak. 2020;20(1):310.
https://doi.org/10.1186/s12911-020-01332-6.

Noorbakhsh-Sabet N, Zand R, Zhang Y, Abedi V. Artificial intelligence transforms the future of healthcare. Am J Med. 2019;132(7):795-801.
https://doi.org/10.1016/j.amjmed.2019.01.017.

Pinsky MR, Bedoya A, Bihorac A, Celi L, Churpek M, Economou-Zavlanos NJ, et al. Use of artificial intelligence in critical care: opportunities and obstacles. Crit Care. 2024;28(1):113.
https://doi.org/10.1186/s13054-024-04860-z.

Carini C, Seyhan AA. Tribulations and future opportunities for artificial intelligence in precision medicine. J Transl Med. 2024;22(1):411.
https://doi.org/10.1186/s12967-024-05067-0.

Tschandl P, Rinner C, Apalla Z, Argenziano G, Codella N, Halpern A, et al. Human-computer collaboration for skin cancer recognition. Nat Med. 2020;26(8):1229-34.
https://doi.org/10.1038/s41591-020-0942-0.

Liu J, Wang C, Liu S. Utility of ChatGPT in clinical practice. J Med Internet Res. 2023;25:e48568.
https://doi.org/10.2196/48568.

Tsoi K, Yiu K, Lee H, Cheng HM, Wang TD, Tay JC, et al. Applications of artificial intelligence for hypertension management. J Clin Hypertens (Greenwich). 2021;23(3):568-74.
https://doi.org/10.1111/jch.14180.

Russell RG, Lovett Novak L, Patel M, Garvey KV, Craig KJT, Jackson GP, et al. Competencies for the use of artificial intelligence-based tools by healthcare professionals. Acad Med. 2023;98(3):348-56.
https://doi.org/10.1097/ACM.0000000000004963.

Ghassemi M, Oakden-Rayner L, Beam AL. The false hope of current approaches to explainable artificial intelligence in healthcare. Lancet Digit Health. 2021;3(11):e745-e750.
https://doi.org/10.1016/S2589-7500(21)00208-9.

Jayakumar P, Moore MG, Furlough KA, Uhler LM, Andrawis JP, Koenig KM, et al. Comparison of an artificial intelligence-enabled patient decision aid vs educational material on decision quality, shared decision-making, patient experience, and functional outcomes in adults with knee osteoarthritis: a randomized clinical trial. JAMA Netw Open. 2021;4(2):e2037107.
https://doi.org/10.1001/jamanetworkopen.2020.37107.

Van Veen D, Van Uden C, Blankemeier L, Delbrouck JB, Aali A, Bluethgen C, et al. Adapted large language models can outperform medical experts in clinical text summarization. Nat Med. 2024;30(4):1134-42.
https://doi.org/10.1038/s41591-024-02855-5.

Hernandez-Boussard T, Bozkurt S, Ioannidis JPA, Shah NH. MINIMAR (minimum information for medical AI reporting): developing reporting standards for artificial intelligence in healthcare. J Am Med Inform Assoc. 2020;27(12):2011-5.

AlSaad R, Abd-Alrazaq A, Boughorbel S, Ahmed A, Renault MA, Damseh R, et al. Multimodal large language models in healthcare: applications, challenges, and future outlook. J Med Internet Res. 2024;26:e59505.
https://doi.org/10.2196/59505.

Kernberg A, Gold JA, Mohan V. Using ChatGPT-4 to create structured medical notes from audio recordings of physician-patient encounters: comparative study. J Med Internet Res. 2024;26:e54419.
https://doi.org/10.2196/54419.

Reddy S, Fox J, Purohit MP. Artificial intelligence-enabled healthcare delivery. J R Soc Med. 2019;112(1):22-8.
https://doi.org/10.1177/0141076818815510.

Lu Y, Melnick ER, Krumholz HM. Clinical decision support in cardiovascular medicine. BMJ. 2022;377:e059818.
https://doi.org/10.1136/bmj-2020-059818.

Hogg HDJ, Martindale APL, Liu X, Denniston AK. Clinical evaluation of artificial intelligence-enabled interventions. Invest Ophthalmol Vis Sci. 2024;65(10):10.
https://doi.org/10.1167/iovs.65.10.10.

Anto A, Sousa-Pinto B, Bousquet J. Anaphylaxis and digital medicine. Curr Opin Allergy Clin Immunol. 2021;21(5):448-54.
https://doi.org/10.1097/ACI.0000000000000764.

Dias R, Torkamani A. Artificial intelligence in clinical and genomic diagnostics. Genome Med. 2019;11(1):70.
https://doi.org/10.1186/s13073-019-0689-8.

Akay EMZ, Hilbert A, Carlisle BG, Madai VI, Mutke MA, Frey D. Artificial intelligence for clinical decision support in acute ischemic stroke: a systematic review. Stroke. 2023;54(6):1505-16. https://doi.org/10.1161/STROKEAHA.122.041442.

Loh HW, Ooi CP, Seoni S, Barua PD, Molinari F, Acharya UR. Application of explainable artificial intelligence for healthcare: a systematic review of the last decade (2011-2022). Comput Methods Programs Biomed. 2022;226:107161. https://doi.org/10.1016/j.cmpb.2022.107161.