

ORIGINAL RESEARCH

Open access

Large Language Model for Automated Summarization of Interdisciplinary Care Team Discussions Using Secure Clinical Meeting Transcripts, Active Problem Lists, Medication Changes, and Discharge Planning Notes

Yousef Al-Qahtani^{1*}, Fahad Al-Salem¹, Abdullah Al-Harbi²

Abstract

Interdisciplinary rounds, discharge planning meetings, and tumor boards contain high-value clinical reasoning that is often only partially reflected in the medical record. These discussions shape treatment priorities, medication decisions, consult plans, and discharge readiness, yet their verbal and collaborative nature makes them difficult to document comprehensively. Manual summarization of care team discussions requires time, attention, and clinical synthesis that busy clinicians may not have during or immediately after meetings. Existing documentation practices often capture final decisions but omit uncertainty, rationale, task ownership, and evolving care coordination needs. This article proposes a large language model pipeline that could summarize interdisciplinary care discussions using secure meeting transcripts combined with active problem lists, medication lists, and discharge planning notes. The objective is to describe a conceptual architecture for generating accurate, structured, and clinically reviewable summaries of team communication. The proposed approach uses retrieval-augmented generation to ground the language model in structured clinical context while processing a diarized transcript of the care discussion. The model would focus on identifying decisions, medication changes, unresolved issues, discharge barriers, and action items requiring follow-up. Conceptually, the pipeline would generate a note-ready summary with lower hallucination risk because the model is constrained by structured clinical anchors and transcript evidence. It could help distinguish new decisions from repeated background information and convert a documentation-light meeting into a structured clinical artifact. A secure, grounded large language model system could support safer and more complete documentation of interdisciplinary care discussions. By combining transcript evidence with patient-specific structured data, such a system could reduce cognitive burden and improve continuity across clinical teams.

Keywords Large language models, Discharge planning, Clinical summarization, Meeting transcripts, Interdisciplinary care, Factual consistency

*Correspondence:

Yousef Al-Qahtani

yousef.qahtani@gmail.com

¹ Department of Digital Health Engineering, Faculty of Medicine, King Saud University, Riyadh, Saudi Arabia

² Department of Clinical Informatics and AI Systems, Faculty of Engineering, King Abdulaziz University, Jeddah, Saudi Arabia

Introduction

Interdisciplinary care team discussions are central to safe inpatient care because they bring together physicians, nurses, pharmacists, case managers, therapists, and specialists to reconcile competing priorities and coordinate next steps.

Tumor boards, rounds, and discharge planning meetings frequently contain nuanced reasoning about clinical trajectory, risks, treatment preferences, and operational constraints that may not be fully captured in standard notes. Prior work on multidisciplinary team meetings and clinical decision support suggests that artificial intelligence could help organize these complex conversations, but the clinical record still often reflects only the final plan rather than the reasoning process that produced it [1-3]. A dedicated summarization system would therefore address a communication gap at the intersection of documentation quality, team cognition, and patient safety.

Current documentation workflows rely on handwritten notes, post-meeting dictation, copied forward plans, or fragmented updates entered by different team members. These practices can omit medication rationale, discharge barriers, responsibility assignments, and unresolved concerns, particularly when discussions involve multiple speakers and rapidly evolving information. Clinical summarization research has shown that hospital course narratives, discharge summaries, and medical notes require synthesis across long, heterogeneous sources rather than simple extraction of isolated facts [3-6]. For interdisciplinary discussions, the challenge is greater because spoken exchanges must be converted into a coherent summary while preserving the distinction between confirmed decisions, tentative plans, and background context.

Large language models create an opportunity to rethink this workflow because they can generate fluent summaries, follow structured prompts, and integrate retrieved patient-specific context into a coherent output. Recent studies on clinical summarization, radiology impression generation, and medical note generation show that transformer-based and large language model approaches can be adapted to clinical language, although they require careful grounding and human evaluation [7-10]. Domain-specific language models and clinical pretraining further indicate that medical vocabulary, abbreviations, and documentation conventions benefit from specialized adaptation rather than generic language modeling alone [11-13]. This makes the present moment suitable for a purpose-built system that summarizes team discussions while remaining constrained by clinical evidence.

The thesis of this EAI article is that a secure, structured large language model pipeline could ingest real-time meeting transcripts together with active problem lists, medication changes, and discharge planning notes to produce a verified care-team summary. The system would not replace clinician judgment; instead, it would generate a reviewable draft that surfaces decisions, pending tasks, and areas of disagreement while linking each claim to transcript or structured data sources. Work on factual consistency, radiology report generation, and privacy-preserving clinical language models supports the need for explicit safeguards against hallucination, unsupported claims, and unauthorized data exposure. A clinically responsible implementation would therefore combine retrieval, source attribution, post-generation verification, and clinician approval before any summary becomes part of the electronic health record.

Background

The role of interdisciplinary care meetings

Interdisciplinary care meetings include intensive care rounds, oncology tumor boards, discharge planning conferences, complex case reviews, and specialty coordination meetings. These settings are information dense because they integrate disease status, treatment response, medication decisions, social needs, patient preferences, bed availability, and follow-up logistics into a shared plan. Studies involving tumor boards and multidisciplinary clinical workflows show that natural language processing and conversational artificial intelligence may support decision organization, but they also reveal the complexity of aligning recommendations with guidelines and local practice constraints [1, 2, 14, 15]. Documentation deficits arise when the final note records only selected conclusions while losing the spoken rationale, role-specific contributions, and conditional plans discussed by the team.

Clinical summarization and documentation

Earlier clinical summarization research has focused largely on written sources such as progress notes, inpatient records, discharge summaries, hospital course sections, radiology reports, and doctor-patient conversations. These studies show that summarization in healthcare requires more than compression: the model must preserve clinical chronology, distinguish active from resolved problems, and maintain consistency with source evidence [4-6, 16]. The shift toward large language models and conversational note generation has broadened the field from document summarization to ambient clinical intelligence, where spoken interactions can be converted into structured documentation [9, 16, 17]. An interdisciplinary meeting summarizer would extend this trajectory from dyadic encounters and written notes to multi-participant care coordination.

Large language models in medicine

Large language models in medicine can support summarization, question answering, documentation drafting, and decision support, but their outputs must be treated as probabilistic drafts rather than authoritative clinical facts. Reviews of large language models in medicine emphasize both their potential and their risks, including hallucination, outdated knowledge, uncalibrated confidence, and variable performance across clinical contexts [18]. Clinical summarization studies similarly indicate that model adaptation and domain-specific prompting may improve usability, yet safety requires independent verification and clinician oversight [7, 19]. For team-discussion summarization, these limitations are especially important because a fluent but unsupported statement could propagate into handoffs, medication plans, or discharge instructions.

Multi-source and grounded summarization

Multi-source grounded summarization is essential when a clinical meeting references information distributed across transcripts, problem lists, medication records, orders, imaging reports, and discharge planning notes. Work on hospital-course summarization demonstrates the need to synthesize longitudinal clinical information rather than merely summarize a single note, while radiology and clinical entity extraction studies show how structured concepts can support factual anchoring [4, 5, 20]. A retrieval-augmented approach could use active problems, medications, and discharge tasks as grounding anchors so that generated claims remain tied to patient-specific evidence. This design would be expected to reduce unsupported content by requiring the model to retrieve relevant structured context before drafting the final summary.

Operational benefits of summarization for team communication

A high-quality team-discussion summary could reduce information loss by preserving decisions, task ownership, discharge barriers, and follow-up needs in a single shared artifact. Work on emergency medicine handoff notes, automatic note generation, and multidisciplinary meeting support suggests that summarization tools may help clinicians manage transitions of care and preserve context across teams [1, 8, 21]. In discharge planning, a grounded summary could highlight whether a barrier is clinical, logistical, medication-related, or social, allowing team members to coordinate around the same representation of the plan. The operational value would therefore come not only from reducing documentation burden but also from improving continuity between rounds, handoffs, consults, and discharge workflows.

Summarization Pipeline Overview

High-level architecture

The proposed architecture begins with secure audio capture and speech-to-text transcription, followed by speaker diarization that separates contributions from attending physicians, residents, nurses, pharmacists, case managers, and other participants. The diarized transcript would be paired with an electronic health record pull of active problems, medication lists, recent medication changes, discharge planning notes, and relevant orders, after which a retrieval-augmented language model would generate a structured summary. A post-processing verifier would compare each generated claim against transcript segments and structured data elements before the draft is presented for clinician

review. This architecture draws conceptually from ambient clinical intelligence, clinical note generation, and hospital-course summarization work, while adding safeguards for multi-participant team communication [8, 9, 16].

Figure 1 illustrates the proposed secure, grounded LLM pipeline for transforming interdisciplinary care team discussions into structured, source-linked, clinician-reviewable documentation.

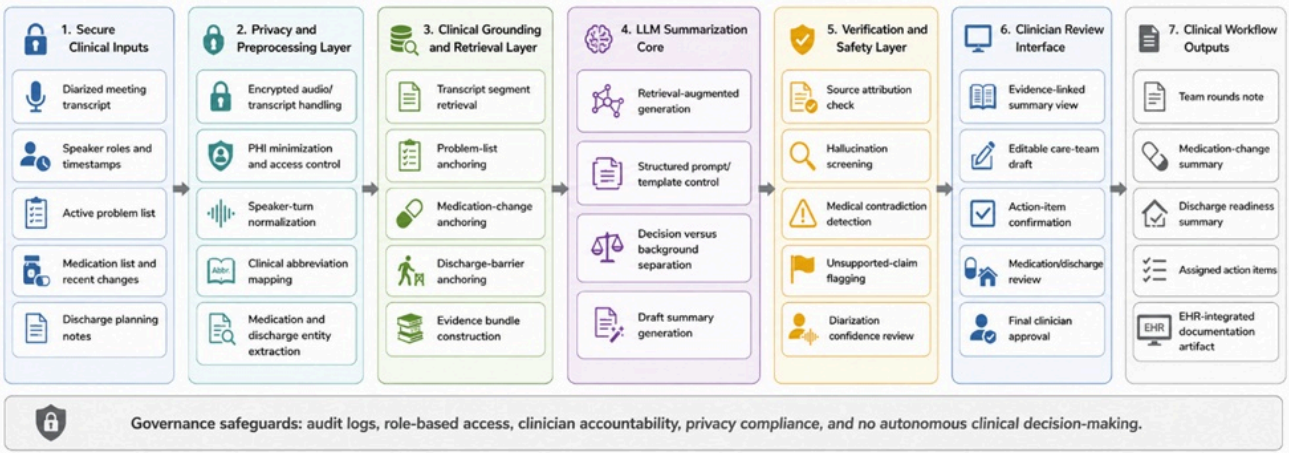


Figure 1. Secure Grounded LLM Architecture for Summarizing Interdisciplinary Care Team Discussions

Core input components

The core inputs would include a timestamped, diarized meeting transcript; an active problem list mapped to clinical terminology; a current medication list with recent starts, stops, dose changes, and holds; and discharge planning notes that identify barriers, responsibilities, and pending milestones. The transcript would provide the conversational evidence, while structured data would supply stable anchors for diagnoses, therapies, and disposition plans. Prior clinical summarization studies show that reliable summaries require awareness of patient-specific context, while entity and relation extraction work in radiology illustrates how structured clinical concepts can support downstream verification [4, 20, 22]. In this design, the transcript and structured records would be treated as complementary evidence sources rather than interchangeable text fields.

Design principles

The design should be safety-first, privacy-preserving, clinician-in-the-loop, and embedded within the normal rounding or meeting workflow. Every generated statement should be traceable to a transcript timestamp or structured data element, and unsupported statements should be withheld or marked for review. Privacy-preserving deployment may involve on-premise inference, encrypted processing, access controls, and data minimization to reduce exposure of protected health information [23, 24]. Clinician review remains central because the model's role is to draft, organize, and surface evidence, not to independently determine the official clinical plan.

Table 1 defines the safety-critical design logic required to convert multidisciplinary discussion into clinically reviewable documentation without allowing unsupported model generation to enter the record.

Table 1. Safety-Critical Design Logic for a Grounded LLM Summarization System in Interdisciplinary Care

Design domain	Core design requirement	Clinical risk addressed	Required system mechanism	Expected contribution to documentation quality

Transcript integrity	Preserve speaker turns, timestamps, and uncertainty markers	Misattributed decisions or loss of qualifications	Role-aware diarization, timestamped transcript segmentation, confidence scoring	Maintains accountability and allows reviewers to trace who said what
Clinical grounding	Link generated claims to active problems, medications, and discharge notes	Hallucinated or context-free summaries	Retrieval-augmented generation using patient-specific structured anchors	Ensures summaries reflect the current clinical situation rather than generic medical knowledge
Medication safety	Distinguish medication starts, stops, holds, dose changes, and reconciliation issues	Unsafe medication misstatement or omitted rationale	Medication-change extraction and structured medication-record comparison	Improves visibility of treatment changes and reduces ambiguity during handoffs
Discharge coordination	Identify barriers, responsibilities, and pending milestones	Lost follow-up tasks or unclear discharge readiness	Discharge-barrier extraction, task-role mapping, and action-item routing	Converts verbal planning into accountable workflow artifacts
Factual consistency	Require every summary statement to be source-supported	Unsupported conclusions entering the EHR	Source attribution, contradiction detection, unsupported-claim flagging	Reduces hallucination risk and improves clinician trust
Human oversight	Require clinician review before EHR integration	Automation bias and premature acceptance of draft content	Editable review interface, evidence links, approval gate	Keeps final responsibility with the clinical team
Privacy governance	Minimize exposure of protected health information	Unauthorized data reuse or disclosure	Encryption, role-based access, data minimization, audit logging	Supports responsible deployment in high-stakes clinical environments

Data Sources and Preprocessing

Secure transcript acquisition and speaker diarization

Secure transcript acquisition would rely on encrypted microphones or approved conferencing infrastructure, role-aware speaker diarization, and clinical speech-to-text systems configured for protected health information. The preprocessing layer would label speakers by role when possible, preserving accountability without unnecessarily exposing identity beyond clinical need. Ambient clinical intelligence datasets and doctor-patient conversation note-generation research demonstrate that spoken clinical language can be transformed into documentation-oriented text, but interdisciplinary meetings require additional handling because multiple speakers may interrupt, correct, or qualify one another [9, 16, 17]. The resulting transcript should preserve timestamps, speaker turns, and uncertainty markers so that downstream summarization can distinguish confirmed decisions from conversational exploration.

Alignment with active problem and medication lists

Alignment with active problem and medication lists would allow the model to ground spoken phrases such as “heart failure,” “CHF,” or “volume overload” in coded clinical problems and related therapies. This preprocessing step would normalize abbreviations, identify synonyms, and link discussion fragments to active or historical conditions so that the summary reflects the current clinical frame rather than isolated utterances. Domain-specific biomedical and clinical language models provide a foundation for recognizing medical vocabulary, abbreviations, and contextualized concepts in clinical text [11–13]. By grounding conversational statements in structured records, the pipeline could reduce ambiguity and make the summary easier for clinicians to verify.

Extraction of medication changes and discharge planning notes

Medication change extraction would identify spoken signals such as starting, stopping, holding, continuing, increasing, decreasing, or reconciling a medication, then compare those signals with the structured medication record. The same process would apply to discharge planning content, where verbal statements about equipment, home services, placement, transportation, follow-up, or caregiver readiness would be linked to discharge planning notes. Prior work on hospital-course summarization and medical note generation shows that clinically useful summaries must synthesize treatment decisions and care transitions across multiple sources rather than simply restate narrative text [4, 5, 8]. A model designed for interdisciplinary care discussions should therefore treat medication and discharge content as high-priority domains for extraction, verification, and task routing.

Preprocessing for privacy and data minimization

Privacy-oriented preprocessing would remove or mask patient-identifiable details that are not necessary for the summarization task, strip conversational filler, and retain only clinically relevant content for model ingestion. The system should minimize exposure by processing the smallest adequate data slice, enforcing access controls, and preventing reuse of identifiable transcripts outside approved clinical governance. Research on privacy-preserving large language models and federated or distributed approaches indicates that clinical deployment must address data security, model access, and institutional control of sensitive records [23, 24]. Data minimization is therefore not only a compliance measure but also a safety design principle that limits the model's opportunity to generate irrelevant or unauthorized content.

LLM Architecture and Adaptation

Language model choice and fine-tuning strategy

The language model could be a compact, domain-adapted system deployed within a secure hospital environment, with parameter-efficient adaptation applied to de-identified clinical meeting transcripts when governance permits. Adaptation would focus on clinical discourse patterns, role-based utterances, medication-change language, discharge planning terminology, and the distinction between decisions, concerns, and action items. Clinical summarization and domain-specific language model studies suggest that adaptation to medical language and documentation conventions is necessary for safe and useful generation [7, 11, 13, 19]. The model should remain constrained as a drafting tool whose outputs require source support and clinician approval before use.

Retrieval-augmented generation for grounding

Retrieval-augmented generation would retrieve relevant problem list entries, medication records, discharge notes, and transcript segments at generation time so that the summary remains anchored in patient-specific evidence. Rather than asking the model to rely on latent knowledge, the pipeline would require it to compose a summary from retrieved sources that correspond to the current discussion. Work on factual consistency, radiology report correctness, and clinical summarization highlights the importance of grounding generated text in source material to reduce unsupported or contradictory claims [25–27]. In this framework, retrieval is not an optional enhancement but the central safety mechanism that determines what the model is allowed to summarize.

Structured output and template enforcement

The output should be constrained to a structured format that includes clinical impression, care decisions, medication changes, action items with responsible roles, discharge readiness, unresolved questions, and items requiring review. Template enforcement could be implemented through controlled prompting, schema-constrained decoding, or post-generation validation that checks whether each section contains only supported content. Studies of emergency medicine handoff notes, doctor-patient conversation documentation, and clinical note generation suggest that structured outputs are more likely to align with workflow needs than free-form summaries alone [9, 16, 21]. A template also allows the system to separate clinically confirmed decisions from pending issues, thereby reducing the risk that tentative discussion becomes documented as a final plan.

Lightweight local deployment and latency considerations

A lightweight local deployment would be appropriate for hospitals that require strict control over clinical transcripts and structured patient data. The model could be optimized for local inference on hospital-grade servers, with privacy safeguards, audit logs, and role-based access integrated into the clinical environment. Privacy-preserving large language model research and broader work on federated learning with language models support the principle that sensitive healthcare data should be processed under strong institutional governance rather than exposed unnecessarily to external systems [23, 24]. Latency should be addressed as a workflow requirement, because the summary is most useful when it is available for review soon after the meeting while the discussion remains fresh.

Ensuring Factual Accuracy and Safety

Hallucination mitigation through source attribution

Hallucination mitigation should begin by requiring every generated statement to map back to a transcript timestamp, structured problem-list entry, medication record, or discharge planning source. This approach would prevent the model from presenting plausible but unsupported clinical content as fact, which is particularly important when summaries influence handoffs, medication reconciliation, or discharge decisions. Prior work on factual consistency in radiology reporting and clinical summarization shows that fluent language is insufficient unless the generated content remains faithful to source evidence [25-27]. The proposed system should therefore flag unsupported statements for clinician review rather than silently include them in the final note.

Medical contradiction detection and escalation

A post-generation contradiction checker would compare the draft summary against the active problem list, medication record, allergies, and relevant structured orders. For example, if the summary suggested initiation of a medication while the structured record indicated a contraindication, hold order, or unresolved safety concern, the system should escalate that claim for clinician verification. Clinical entity and relation extraction methods could support this verification layer by identifying whether generated statements preserve the relationships present in source records [20, 22]. This checker would not make autonomous safety decisions, but it would help surface inconsistencies that might otherwise be missed in a busy clinical workflow.

Human-in-the-loop verification and correction

Human-in-the-loop verification is essential because the proposed model would function as a clinical documentation assistant rather than an independent decision maker. The attending physician or designated reviewer should be able to inspect evidence links, edit summary language, confirm action items, and approve the note before it enters the electronic health record. Studies of clinical note generation, handoff summarization, and medical large language models emphasize that clinician judgment remains necessary to assess adequacy, safety, and usefulness in context [9, 18, 21]. Logged

corrections could also become governed feedback for later model refinement, provided privacy and institutional review requirements are met.

Interpretability, Verification, and Clinical Review

Traceable summary annotations

Traceable annotations should allow clinicians to click from each generated claim to the corresponding transcript line, medication record, problem-list entry, or discharge planning note. This interface design would make verification faster because reviewers would not need to search manually across the transcript and electronic health record to confirm the basis for a statement. The need for traceability is consistent with factuality-focused work in radiology report generation and clinical summarization, where source grounding helps distinguish supported findings from generated assumptions [25-27]. For team discussions, traceability also helps preserve accountability by showing whether a decision was explicitly stated, inferred from context, or still unresolved.

Clinician feedback loop for continuous improvement

A clinician feedback loop would capture edits, deletions, reordered sections, and rejected claims as signals for future adaptation of the summarization system. These corrections could reveal recurring local patterns, such as how a particular ICU team phrases discharge readiness, how pharmacists discuss medication reconciliation, or how oncology boards frame guideline-based recommendations. Parameter-efficient fine-tuning and domain adaptation approaches suggest that models can be adjusted toward clinical documentation needs without requiring wholesale retraining of a foundation model [11, 13, 19]. Such learning should occur under governance that separates quality improvement from unsafe automatic updating, ensuring that clinician feedback improves reliability without introducing unreviewed behavior.

Verification metrics and trust-building

The system should present simple verification indicators, such as whether each summary section is fully source-linked, partially source-linked, or requires review. These indicators would support appropriate trust by encouraging clinicians to inspect uncertain content instead of treating the generated note as self-validating. Prior clinical summarization and ambient documentation studies show that human evaluation must consider completeness, accuracy, safety, and practical usefulness rather than relying only on automated text-similarity metrics [4, 7, 16]. A useful interface would therefore make model uncertainty visible and actionable, helping clinicians trust the workflow while still maintaining final responsibility for the clinical note.

Integration into Care Team Workflow

Post-meeting summary and EHR integration

After clinician approval, the summary could be posted as a structured “Team Rounds Note” or similar documentation artifact within the electronic health record. The note would ideally include current clinical impression, decisions made, medication changes, discharge readiness, unresolved issues, and assigned action items, making the meeting output visible to all care team members. Work on emergency medicine handoff notes, discharge summarization, and multidisciplinary team meeting support suggests that structured summaries may improve continuity when care responsibility shifts across clinicians and services [1, 4, 21]. Integration into the electronic health record should be designed to reduce duplication rather than add another documentation burden.

Real-time action item alerts

The same pipeline could identify action items during or immediately after a meeting and route them through secure messaging or task management systems to the responsible clinician or team role. Examples could include requesting a consult, confirming a medication reconciliation issue, arranging home services, or clarifying discharge transportation, but each alert should remain reviewable and grounded in the transcript. Prior work on clinical note generation and multidisciplinary care support indicates that summarization has operational value when it transforms conversation into actionable workflow artifacts [1, 8, 15]. Real-time alerts should therefore be treated as an extension of the approved summary, not as independent automated orders.

Evaluation Strategy

Intrinsic summary quality metrics

Intrinsic evaluation should assess whether generated summaries are coherent, complete, source-faithful, and clinically well structured when compared with manually written meeting notes. Automated metrics such as lexical overlap, semantic similarity, and factual consistency measures could be used as supporting signals, but they should not be treated as sufficient evidence of clinical safety. Research on hospital-course summarization, radiology report generation, and factual consistency evaluation shows that automated metrics can help characterize model behavior while still missing clinically important omissions or unsupported claims [4, 5, 10, 27]. The evaluation strategy should therefore use automated measures as screening tools alongside clinician review.

Table 2 provides a clinical readiness evaluation matrix that separates linguistic quality from factual safety, workflow utility, and governance requirements.

Table 2. Evaluation Matrix for Clinical Readiness of Interdisciplinary Meeting Summarization

Evaluation dimension	Primary assessment question	Suggested evidence source	Minimum acceptable standard	Deployment implication
Completeness	Does the summary capture major decisions, medication changes, discharge barriers, unresolved questions, and action items?	Blinded multidisciplinary reviewer scoring	No clinically important category consistently omitted	Determines whether the system can support team documentation
Source faithfulness	Is each generated claim supported by transcript or structured EHR evidence?	Claim-level evidence audit	Unsupported clinical claims must be rare and clearly flagged	Determines hallucination safety
Medication accuracy	Are medication starts, stops, holds, dose changes, and rationales represented correctly?	Pharmacist review against transcript and medication record	No unflagged clinically meaningful medication errors	Determines suitability for medication-sensitive workflows
Discharge planning utility	Does the summary clarify barriers, responsibilities, and readiness status?	Case manager and discharge coordinator review	Action items and barriers must be assignable and verifiable	Determines workflow value beyond documentation burden reduction
Contradiction detection	Are conflicts between generated summary and structured record identified?	Structured-record comparison and clinician adjudication	High-risk contradictions must	Determines safety of EHR integration

			be escalated before approval	
Reviewer burden	Does the draft reduce or increase clinician documentation workload?	Time-motion study and user feedback	Review time must be justified by improved completeness and safety	Determines practical adoption feasibility
Trust calibration	Do clinicians appropriately inspect uncertain or flagged content?	Interface-use logs and qualitative interviews	Users should not accept flagged claims without review	Determines whether the system promotes safe reliance
Governance readiness	Are privacy, audit, access, and accountability controls operational?	Institutional compliance and security review	No deployment without auditable access and approval workflows	Determines institutional readiness for pilot implementation

Clinical adequacy and safety review

Clinical adequacy should be assessed through blinded review by physicians, nurses, pharmacists, case managers, and other team members who understand the meeting context. Reviewers should judge whether the summary captures the major decisions, medication changes, discharge barriers, unresolved questions, and responsibilities without adding unsupported or unsafe content. Prior work on large language model clinical summarization and ambient clinical intelligence emphasizes the importance of human evaluation for adequacy, completeness, clinical usefulness, and safety [7, 9, 16]. This review process should also distinguish minor wording edits from clinically meaningful corrections, because the latter carry greater implications for deployment readiness.

Prospective workflow and efficiency impact

Prospective evaluation should examine whether the system improves documentation completeness, reduces duplicated documentation work, supports handoffs, and improves team awareness of pending tasks. It could also assess clinician satisfaction, perceived cognitive burden, timeliness of documentation, and whether discharge-related issues are surfaced earlier in the workflow. Studies involving handoff notes, automated clinical note generation, and multidisciplinary meeting support suggest that workflow impact should be evaluated in the setting where the tool is actually used rather than inferred solely from offline summary quality [1, 8, 21]. The strongest evaluation would therefore combine intrinsic summary review with prospective observation of team communication and documentation processes.

Limitations

Speech recognition and diarization errors

The pipeline would depend on accurate audio capture, transcription, and speaker diarization, all of which may be challenged by noisy rooms, overlapping speech, accents, interruptions, and role changes during meetings. If the transcript misattributes a medication recommendation to the wrong speaker or omits a qualification, the generated summary could propagate an error unless the verification interface makes the source uncertainty visible. Work on ambient clinical intelligence and conversation-based note generation shows that spoken clinical documentation requires careful handling of dialogue structure before it can be converted into reliable notes [9, 16, 17]. For this reason, transcript quality and diarization confidence should be treated as safety-relevant inputs rather than purely technical preprocessing details.

Clinical context and trust

A language model may miss nonverbal cues, institutional context, team hierarchy, or subtle uncertainty conveyed through tone and timing. Clinicians may also overtrust a polished generated summary or undertrust the system because of concern about hallucination, privacy, or workflow disruption. Reviews of large language models in medicine emphasize that clinical deployment must account for limitations in reasoning, factuality, transparency, and user trust [18, 23, 27]. Adoption would therefore require careful governance, training, auditability, and a culture in which the model is understood as a documentation support tool rather than a substitute for professional judgment.

Conclusion

The proposed large language model pipeline would transform interdisciplinary care discussions into structured, reviewable clinical summaries. By combining secure meeting transcripts with active problem lists, medication changes, and discharge planning notes, the system could preserve decisions, rationale, unresolved questions, and action items that are often lost after rounds or tumor boards.

Its key strengths are real-time capture, grounded generation from structured patient data, source traceability, and integration into documentation and task workflows. These features would allow the system to support team communication while preserving clinician oversight and accountability.

Important challenges remain, including speech recognition errors, diarization uncertainty, hallucination risk, privacy governance, and cultural adoption in high-stakes clinical environments. The system would need to be evaluated carefully as a safety-sensitive communication tool rather than simply as a generic summarization model.

Future pilot implementations in intensive care units, oncology tumor boards, and discharge planning meetings should examine safety, efficiency, clinician workload, and communication outcomes. Such pilots would help determine whether secure, grounded summarization can become a reliable part of clinical team documentation.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 22 Nov 2023

Revised: 28 Dec 2023

Accepted: 05 Feb 2024

Published online: 25 February 2024

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at <https://creativecommons.org/licenses/by-nc-sa/4.0/>.

References

- Ali SR, Dobbs TD, Tarafdar A, Strafford H, Fonferko-Shadrach B, Sinha S, et al. Natural language processing to automate a web-based model of care and modernize skin cancer multidisciplinary team meetings. *Br J Surg*. 2024;111(1):znad347.
- Sorin V, Klang E, Sklair-Levy M, Cohen I, Zippel DB, Balint Lahat N, et al. Large language model (ChatGPT) as a support tool for breast tumor board. *NPJ Breast Cancer*. 2023;9(1):44.
<https://doi.org/10.1038/s41523-023-00557-8>.
- Gabriel J, Gabriel A, Shafik L, Alanbuki A, Larner T. AI in the urology MDM: can ChatGPT suggest EAU guideline-recommended prostate cancer treatments? *BJU Int*. 2024;133(3):259-61.
<https://doi.org/10.1111/bju.16290>.
- Searle T, Ibrahim Z, Teo J, Dobson RJ. Discharge summary hospital course summarisation of inpatient electronic health record text with clinical concept guided deep pre-trained transformer models. *J Biomed Inform*. 2023;141:104358.
<https://doi.org/10.1016/j.jbi.2023.104358>.
- Adams G, Alsentzer E, Ketenci M, Zucker J, Elhadad N. What's in a summary? Laying the groundwork for advances in hospital-course summarization. In: *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Stroudsburg (PA): Association for Computational Linguistics; 2021. p. 4794-811.
- Ando K, Okumura T, Komachi M, Horiguchi H, Matsumoto Y. Is artificial intelligence capable of generating hospital discharge summaries from inpatient records? *PLOS Digit Health*. 2022;1(12):e0000158.
<https://doi.org/10.1371/journal.pdig.0000158>.
- 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>.
- Yuan D, Rastogi E, Naik G, Rajagopal SP, Goyal S, Zhao F, et al. A continued pretrained LLM approach for automatic medical note generation. In: *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 2: Short Papers)*. Stroudsburg (PA): Association for Computational Linguistics; 2024. p. 565-71.
- Giorgi J, Toma A, Xie R, Chen S, An K, Zheng G, et al. WangLab at MEDIQA-Chat 2023: clinical note generation from doctor-patient conversations using large language models. In: *Proceedings of the 5th Clinical Natural Language Processing Workshop*. Stroudsburg (PA): Association for Computational Linguistics; 2023. p. 323-34.
- Sun Z, Ong H, Kennedy P, Tang L, Chen S, Elias J, et al. Evaluating GPT4 on Impressions Generation in Radiology Reports. *Radiology*. 2023;307(5):e231259.
<https://doi.org/10.1148/radiol.231259>.
- Alsentzer E, Murphy J, Boag W, Weng WH, Jindi D, Naumann T, et al. Publicly available clinical BERT embeddings. In: *Proceedings of the 2nd Clinical Natural Language Processing Workshop*. Stroudsburg (PA): Association for Computational Linguistics; 2019. p. 72-8.

Lee J, Yoon W, Kim S, Kim D, Kim S, So CH, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*. 2020;36(4):1234-40.

Gu Y, Tinn R, Cheng H, Lucas M, Usuyama N, Liu X, et al. Domain-specific language model pretraining for biomedical natural language processing. *ACM Trans Comput Healthc*. 2021;3(1):1-23.
<https://doi.org/10.1145/3458754>.

Ansoborlo M, Gaborit C, Grammatico-Guillon L, Cuggia M, Bouzille G. Prescreening in oncology trials using medical records: natural language processing applied on lung cancer multidisciplinary team meeting reports. *Health Informatics J*. 2023;29(1):14604582221146709.
<https://doi.org/10.1177/14604582221146709>.

Choo JM, Ryu HS, Kim JS, Cheong JY, Baek SJ, Kwak JM, et al. Conversational artificial intelligence (chatGPT™) in the management of complex colorectal cancer patients: early experience. *ANZ J Surg*. 2024;94(3):356-61.
<https://doi.org/10.1111/ans.18749>.

Yim WW, Fu Y, Ben Abacha A, Snider N, Lin T, Yetisgen M. ACI-Bench: a novel ambient clinical intelligence dataset for benchmarking automatic visit note generation. *Sci Data*. 2023;10(1):586.
<https://doi.org/10.1038/s41597-023-02487-3>.

Wang J, Yao Z, Yang Z, Zhou H, Li R, Wang X, et al. NoteChat: a dataset of synthetic patient-physician conversations conditioned on clinical notes. In: Findings of the Association for Computational Linguistics: ACL 2024. Stroudsburg (PA): Association for Computational Linguistics; 2024. p. 15183-201.

Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. *Nat Med*. 2023;29(8):1930-40.
<https://doi.org/10.1038/s41591-023-02448-8>.

Goswami J, Prajapati KK, Saha A, Saha AK. Parameter-efficient fine-tuning large language model approach for hospital discharge paper summarization. *Appl Soft Comput*. 2024;157:111531.
<https://doi.org/10.1016/j.asoc.2024.111531>.

Jain S, Agrawal A, Saporta A, Truong SQ, Duong DN, Bui T, et al. RadGraph: extracting clinical entities and relations from radiology reports. *arXiv [Preprint]*. 2021:arXiv:2106.14463.

Hartman V, Zhang X, Poddar R, McCarty M, Fortenko A, Sholle E, et al. Developing and evaluating large language model-generated emergency medicine handoff notes. *JAMA Netw Open*. 2024;7(12):e2448723.
<https://doi.org/10.1001/jamanetworkopen.2024.48723>.

Delbrouck JB, Chambon P, Chen Z, Varma M, Johnston A, Blankemeier L, et al. RadGraph-XL: a large-scale expert-annotated dataset for entity and relation extraction from radiology reports. In: Findings of the Association for Computational Linguistics: ACL 2024. Stroudsburg (PA): Association for Computational Linguistics; 2024. p. 12902-15.

Wiest IC, Ferber D, Zhu J, van Treeck M, Meyer SK, Juglan R, et al. Privacy-preserving large language models for structured medical information retrieval. *NPJ Digit Med*. 2024;7(1):257.
<https://doi.org/10.1038/s41746-024-01233-2>.

Chen C, Feng X, Li Y, Lyu L, Zhou J, Zheng X, et al. Integration of large language models and federated learning. *Patterns (N Y)*. 2024;5(12):101098.
<https://doi.org/10.1016/j.patter.2024.101098>.

Miura Y, Zhang Y, Tsai E, Langlotz C, Jurafsky D. Improving factual completeness and consistency of image-to-text radiology report generation. In: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg (PA): Association for Computational Linguistics; 2021. p. 5288-304.

Delbrouck JB, Chambon P, Bluethgen C, Tsai E, Almusa O, Langlotz C. Improving the factual correctness of radiology report generation with semantic rewards. In: Findings of the Association for Computational Linguistics: EMNLP 2022. Stroudsburg (PA): Association for Computational Linguistics; 2022. p. 4348-60.

Luo Z, Xie Q, Ananiadou S. Factual consistency evaluation of summarization in the era of large language models. *Expert Syst Appl.* 2024;254:124456.
<https://doi.org/10.1016/j.eswa.2024.124456>.