ClinBridge: An AI-Assisted Framework for Structured Clinical Referral and Handover Using Intelligent Information Extraction and SBAR Generation

Prasad Gadge[1], Tejaswini KoliZ[2], Amey Thite[3], Arvind Rebari[4], Ashwini Sanjay Gagare[5]

[1,2,3,4,5]*Department of Science & Computer Science, MIT Arts, Commerce and Science College, Alandi (D), Pune, Maharashtra, India*

doi.org/10.64643/ijirt.208607-459

***Abstract*—Clinical referral and handover communication is often unstructured, incomplete, duplicated, or internally inconsistent, forcing the receiving clinician to manually reconstruct the patient story before any decision can be made.**

This paper presents ClinBridge, a proposed AI-assisted framework for structured detection of referral-quality problems and generation of a clinician-reviewed handover. The system accepts an incoming referral document, extracts clinical information using a large language model constrained to a strict schema, and passes the result through a deterministic validation layer that checks completeness, flags cross-document inconsistencies, and computes an explainable Referral Quality Score. Instead of returning only a summarized narrative, the proposed system stores field-level evidence source text, page, and confidence and produces a missing-information label, a conflict label where relevant, and a Situation-Background-Assessment-Recommendation (SBAR) draft in which every sentence is marked as documented fact, extracted value, or AI-generated inference.

The paper reviews existing research on structured handover communication, LLM-based clinical extraction, and FHIR/ABDM interoperability, describes the proposed architecture and workflow, and discusses an evaluation framework using field-level precision, recall, F1-score, conflict-detection accuracy, and workflow-usefulness metrics. The work is intended as a practical research direction for referral-quality screening that keeps the clinician as the final decision-maker, without assuming or claiming autonomous diagnostic or treatment-decision capability.

Index Terms*—Clinical Handover, Deterministic Validation, FHIR, Healthcare AI, Human-in-the-Loop, Information Extraction, Large Language Models, Referral Quality, SBAR, Responsible AI.

I. INTRODUCTION

Referral is the mechanism by which responsibility for a patient's care is transferred between clinicians, departments, or facilities. Referral documents combine typed notes, handwritten entries, scanned reports, and free-text summaries, and the information they contain is often scattered, duplicated, or contradictory across pages. The receiving clinician must read the entire document, mentally reconstruct the patient's condition, identify what is missing, and, when something is unclear, contact the referring side directly. Earlier work established the clinical importance of structured handover communication, while later evidence syntheses examined how consistently tools such as SBAR are actually implemented and what effect they have on safety outcomes [1], [2].

The problem is not simply that a referral is long or informally written. A short, complete referral can be perfectly usable. The concern begins when required information is silently absent, when the same fact is stated differently in two places, or when a generated summary blurs the line between what was documented and what was inferred. Regulatory and safety literature on care transitions has repeatedly linked such gaps to delayed investigations, delayed treatment, and avoidable errors [3]. For a receiving clinician, identifying these gaps manually is difficult because referrals are heterogeneous, multi-page, and frequently mix structured fields with narrative text.

This paper proposes ClinBridge, an automated referral-quality and handover framework designed around four goals: (1) extract relevant clinical information from a given referral document, (2) identify which expected fields are missing or documented as explicitly negative, (3) detect inconsistencies using deterministic rules supported by AI-based extraction, and (4) generate a structured SBAR handover in which every statement is traceable to source evidence. The objective is not to diagnose the patient or judge the referring clinician, but to identify documentation gaps that deserve review before the referral reaches the next stage of care.

II. LITERATURE REVIEW

Structured-handover research has developed from descriptive communication-quality studies toward more rigorous evidence synthesis. A 2025 evidence review conducted for Making Healthcare Safer IV found low-certainty evidence that SBAR improves patient-safety clinical outcomes and moderate-certainty evidence for the related I-PASS protocol, while noting that faithful implementation of SBAR in routine practice is difficult to achieve [1]. A systematic review focused specifically on SBAR fidelity reported similar caution, observing that many published studies do not verify whether SBAR was actually used as designed at the point of care [2]. In surgical handover, structured tools have been associated with fewer omissions and documentation errors, though effect sizes vary considerably by setting [3]. Collectively, this literature supports SBAR as a communication scaffold worth automating, but argues against overstating any guaranteed effect on patient outcomes. Separately, clinical natural-language-processing research has moved from rule-based and statistical extraction toward large language models capable of few-shot clinical entity extraction [4]. Methodology reviews of clinical concept extraction describe this as a long-running shift in technique rather than a solved problem [5]: generative extraction is fluent but not inherently faithful to the source document, and can produce structured output that looks correct while containing inferred or fabricated values.

This motivates treating extraction and validation as separate layers rather than trusting a single model end-to-end.

On interoperability, HL7 FHIR provides a resource-based model for representing patients, requests, observations, and clinical documents, and in India the FHIR Implementation Guide for the Ayushman Bharat Digital Mission (ABDM) constrains FHIR R4 with India-specific identifier and document profiles [6], operating within the national digital-health ecosystem led by the National Health Authority [7]. FHIR's ServiceRequest resource is defined for exactly this kind of referral/service request and carries minimum conformance expectations under the ABDM guide [8]. ClinBridge targets alignment with this guide rather than a self-invented schema, so that a later production version does not require a full data-model rewrite. Taken together, the prior work motivates but does not itself deliver a system that combines structured communication (SBAR), faithful extraction (LLM plus provenance), and standards-based interoperability (FHIR/ABDM) inside one referral-quality pipeline which is the contribution of this paper.

III. CHALLENGES IN CLINICAL REFERRAL AND HANDOVER

Referral documents can be described as information-transfer artifacts that fail in observable, recurring ways. The categories below are the ones most relevant to automated detection because they leave evidence in the document's content or structure that a system can check.

3.1. Missing and Incomplete Information

A referral may omit fields that are expected given its own content for example, a medication is named without a dose, or a respiratory complaint is documented without an SpO_2 reading. A detector should not treat every absent field as equally important; expectation should be conditional on what else is documented.

3.2. Cross-Document and Cross-Page Inconsistencies

The same conceptual field patient age, medication dose, allergy status, referral reason can be stated differently across pages or sections of the same referral. A system that silently picks one value hides the discrepancy from the clinician who needs to resolve it.

3.3. Undocumented vs. Explicit Negative Findings

An absent allergy statement is not evidence of "no allergies"; it is evidence that the referral does not say. Conflating these two states is a common and clinically significant failure mode in naive summarization.

3.4. Ambiguous Fact-versus-Inference Boundaries

A generated summary can present an inferred conclusion with the same confidence as a directly documented fact. Without an explicit label distinguishing the two, a receiving clinician cannot tell what the referral actually says from what a model has synthesized.

3.5. Referral Quality Conflated with Clinical Urgency

A poorly documented referral is not necessarily a low-urgency one, and a well-documented referral is not necessarily safe to deprioritize. Any automated quality signal must be kept separate from any notion of clinical risk or triage priority.

Fig. 1. ClinBridge Layered System Architecture

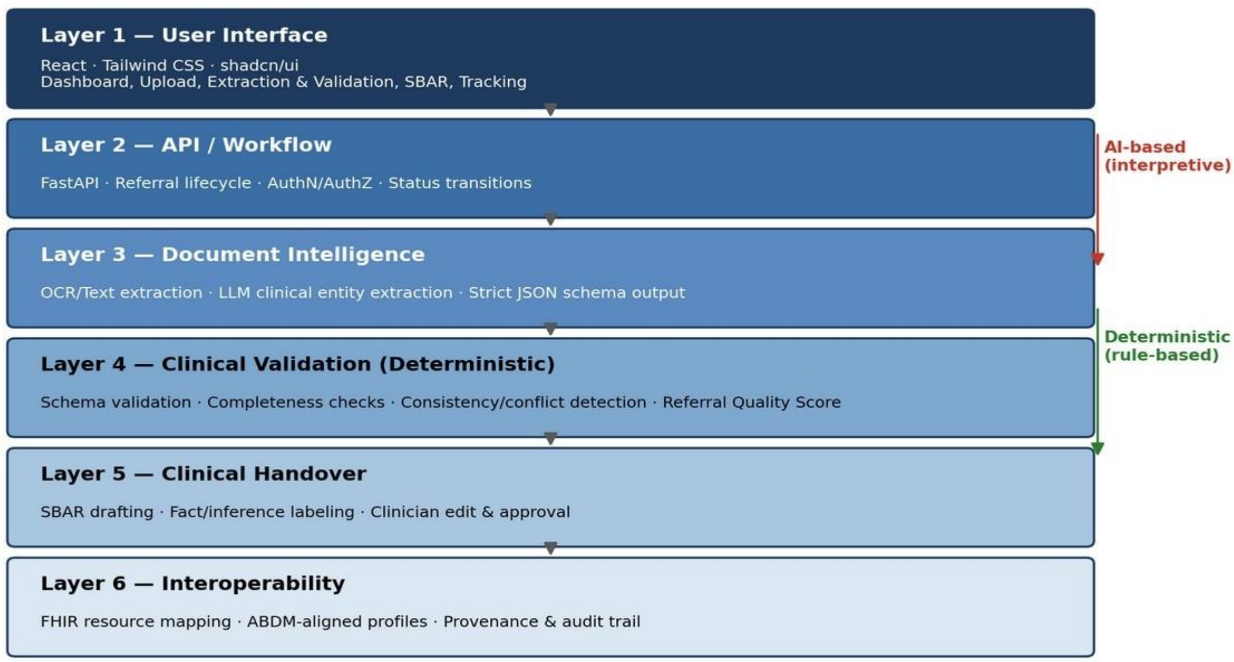

Fig. 1. Proposed layered architecture for AI-assisted clinical referral and handover validation.

IV. PROPOSED SYSTEM

The proposed system is a workflow tool in which a clinician or referral coordinator uploads an incoming referral document. The system extracts visible clinical text and structural information, converts it into a schema-constrained representation, passes it through a deterministic validation layer, and returns a Referral Quality Score together with a clinician-editable SBAR draft. The final output always retains the source evidence behind every field, missing-information flag, and conflict.

Document Ingestion.

The first stage accepts typed PDFs, scanned images, and referral forms. Text extraction is used for typed content and OCR fallback for scanned or image-based referrals, followed by page segmentation before the document reaches the extraction stage. Handwriting recognition and multilingual OCR are treated as post-MVP capabilities because their error rates would otherwise make validation results difficult to interpret.

Feature and Field Extraction.

An LLM reads the pre-processed document and maps its content onto a predefined clinical JSON schema covering patient demographics, referral metadata, history, vitals, diagnosis, medications, allergies, investigations, and the referral request itself. Every extracted field carries a source-text span, source page, and a confidence indicator, and is tagged with a review state documented, not_documented, uncertain, or conflicting so that absence is never silently converted into a negative finding.

Rule-Based Validation.

Some patterns have strong observable rules. A medication with a name but no dose can be flagged as incomplete_medication_information. A respiratory complaint without an SpO_2 reading can be flagged as missing_spo2. Rules provide interpretable evidence and give the downstream scoring stage a stable, auditable basis independent of any single model version.

Deterministic Scoring.

Validated fields feed into an explainable, weighted Referral Quality Score covering core data, referral context, vitals, medications/allergies, and cross-field consistency. Every deduction is itemized (for example, "Missing SpO_2: −8") so the score can be inspected rather than treated as an opaque output.

Evidence and SBAR Drafting.

The validated structure is converted into an SBAR draft in which every sentence is labeled as a documented fact, an extracted value with its source, or an AI-generated inference.

This evidence trail makes the output reviewable rather than an unexplained model summary, and the draft requires explicit clinician edit and approval before it is considered final.

V. EXTRACTION AND VALIDATION WORKFLOW

- Upload Referral: the user submits the source document (PDF/image) for a given patient.
- Render and Extract: OCR/text extraction and page segmentation produce machine-readable content.
- Map to Schema: the LLM extracts clinical fields with source citations and confidence values.
- Run Deterministic Rules: completeness and consistency rules evaluate the extracted structure.
- Compute Quality Score: itemized, weighted score reflecting completeness and consistency.
- Clinician Review: missing fields and conflicts are confirmed, corrected, or resolved.
- Generate SBAR: a fact/inference-labeled draft is produced from the validated structure.
- Approve and Track: the clinician approves the handover, which is then tracked through to the receiving facility.

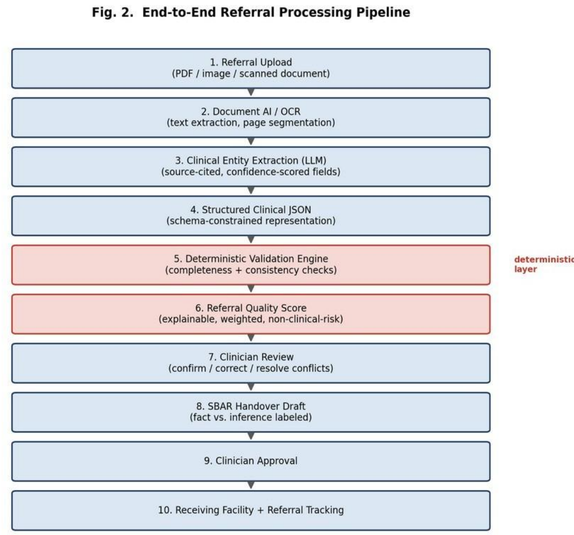

Fig. 2. End-to-end referral processing pipeline from upload to referral tracking.

VI. COMPARISON WITH MANUAL REFERRAL REVIEW

Table I summarizes how the proposed system differs from manual or basic referral review across the dimensions most relevant to referral-quality screening.

TABLE I. COMPARISON WITH MANUAL REFERRAL REVIEW

Aspect	Manual / Basic Review	Proposed ClinBridge System
Scale	Limited to a few referrals at a time	Can process many referrals in parallel
Multi-page documents	Depends on reviewer attention	Systematic page-by-page field extraction
Missing-field detection	Manual reading, easy to overlook	Tiered rule-based completeness checks
Cross-page conflicts	Hard to notice consistently	Explicit, non-silent conflict surfacing
Quality signal	Subjective clinician impression	Itemized, deterministic, explainable score
Interpretability	High for the individual reviewer	Evidence and source citation for every field
Consistency	Depends on reviewer and workload	Same deterministic rules applied every time

VII. DATASET AND SCENARIO DESIGN

Because no real patient data is used at this stage, evaluation depends on a deliberately constructed synthetic referral set with controlled variation, so that extraction and validation performance can be attributed to known, injected defects rather than to unknown properties of an uncontrolled sample. Planned scenarios include a complete baseline referral, a missing allergy statement, a missing medication dose, a missing SpO_2 reading against a respiratory complaint, a conflicting patient age across pages, a conflicting medication dose, duplicate investigation results, degraded OCR quality, and multi-page mixed structured/unstructured content. Each scenario is authored with a ground-truth annotation file recording the exact fields and conflicts intentionally injected, and each is generated in multiple specialty variants general medicine, emergency, and pediatrics to check that rules generalize beyond one referral style.

Fig. 3. Extraction → Validation → SBAR Data Flow with Field-Level States

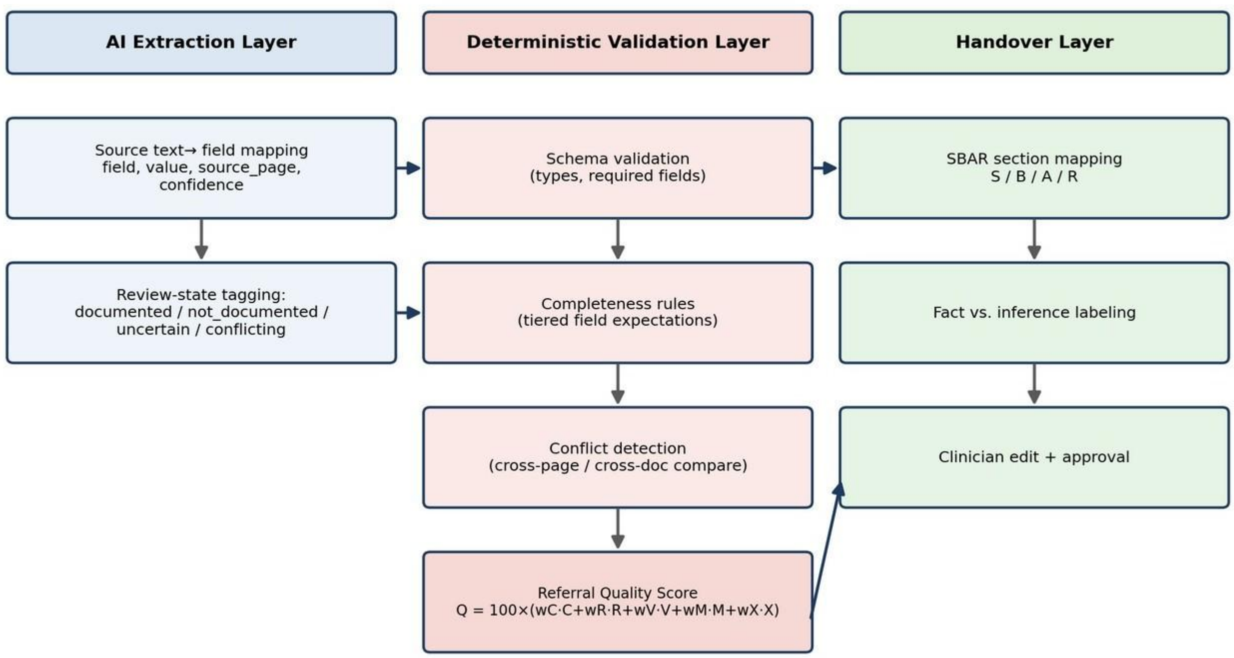

Fig. 3. Field-level data flow from AI extraction through deterministic validation into the SBAR handover.

VIII. EVALUATION METHODOLOGY

The proposed system should be evaluated on a held-out synthetic test set not used during rule or prompt development. Accuracy alone is not sufficient because missing-field and conflict cases are less common than complete, consistent fields. Field-level precision measures how many extracted or flagged items are correct, recall measures how many correct items are found, and F1-score balances the two. Evaluation should additionally include exact-match accuracy for structured values such as dates and numeric fields, source-page attribution accuracy, missing-field detection precision/recall against the injected scenarios, conflict-detection precision/recall, and the false-alert rate on the complete baseline referral, which should be zero.

A second evaluation track measures workflow usefulness rather than raw extraction accuracy: SBAR completeness against the required S/B/A/R sub-elements, the rate of sentences not traceable to a documented source, the accuracy of the fact-versus-inference label attached to each sentence, the edit distance between the drafted and clinician-approved handover, and the time from upload to a review-ready state. Consistent with the cautious framing of the SBAR evidence base [1]– [3], this evaluation is designed to measure documentation quality and workflow efficiency, not patient-outcome improvement; any claim about clinical outcomes would require a separate, prospective clinical study outside the scope of this work.

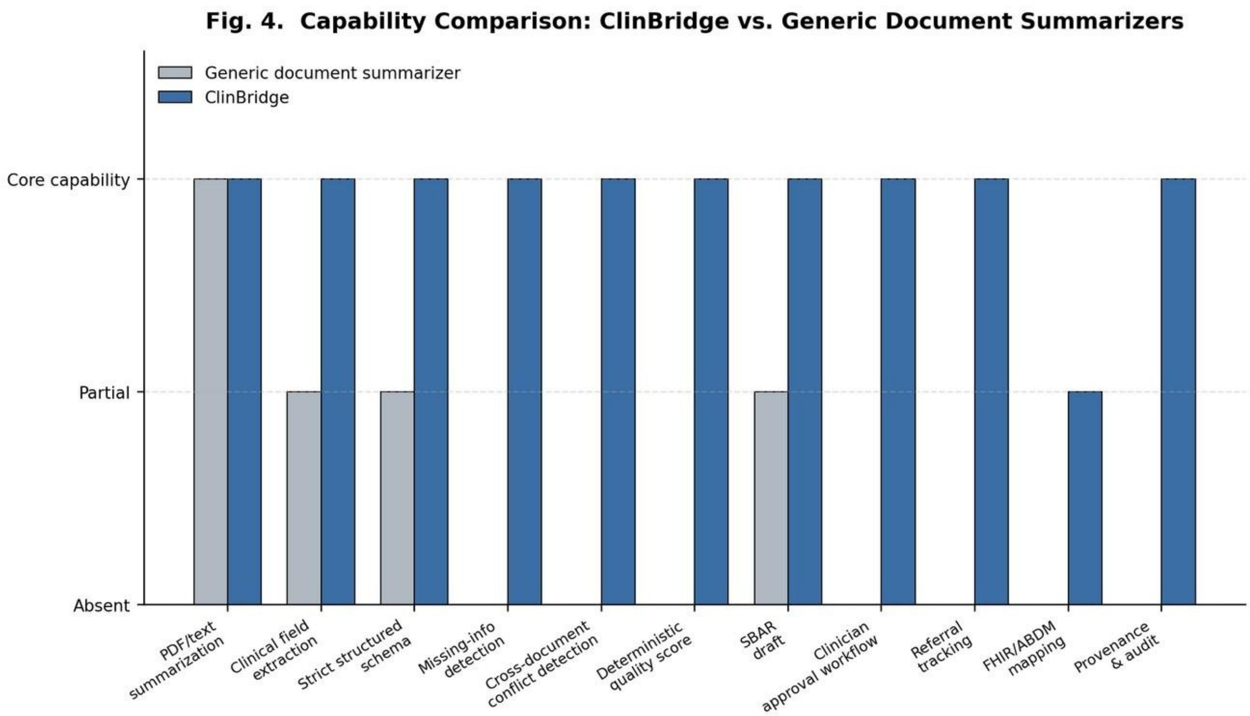

Fig. 4. Capability comparison between ClinBridge and a generic LLM-based document summarizer.

IX. ILLUSTRATIVE CLINICAL REFERRAL CASE

Consider a referral for a patient with a respiratory complaint. The document lists the presenting complaint and a provisional diagnosis but does not include an SpO_2 reading anywhere in the text. The system would not fabricate a value; it records the field as not_documented, raises a missing_spo2 flag tied to the conditional completeness rule, and deducts a fixed, itemized amount from the Referral Quality Score. The clinician reviewing the referral sees exactly which rule fired and why, and can add the missing value or contact the referring facility before approving the handover.

A second example is an age conflict. Suppose page one of the referral states the patient's age as 54 and page three, in a different section, states 57. The system does not silently choose one value; it surfaces both source values with their page numbers, marks the field needs_clinician_review, and withholds a final Referral Quality Score contribution for that field until the conflict is resolved. These examples illustrate the intended behavior on realistic, imperfect referrals rather than a claim that any particular referring facility currently produces inconsistent documentation.

X. BENEFITS OF THE PROPOSED APPROACH

- Scalability: automated extraction can support review of more referrals than a human reviewer can inspect line by line.
- Consistency: the same completeness and consistency rules are applied identically across referrals and reviewers.
- Evidence-based reporting: every flag and score deduction retains the source text or field that produced it.
- Early screening: receiving clinicians can prioritize referrals that most need clarification before acting.
- Adaptability: new completeness or conflict rules can be added as documentation practices evolve.
- Interoperability groundwork: a FHIR/ABDM-aligned structure eases future exchange with receiving health information systems.

XI. CHALLENGES AND LIMITATIONS

Handwritten and low-quality scanned referrals are a first major challenge; OCR error rates on such documents are high enough that the current design excludes them from the MVP rather than risk silently wrong extraction. Context is a second challenge: an

SpO_2 omission may be irrelevant for a non-respiratory complaint, and a low medication dose may be clinically correct rather than incomplete. The system mitigates this with conditional, context-aware rules rather than blanket field requirements, but cannot eliminate the need for clinician judgment.

Dataset and annotation quality are a further limitation: labeling what counts as a clinically meaningful conflict or omission involves human judgement, and different clinician reviewers may disagree, which argues for clear annotation guidelines and multiple reviewers when building the training and evaluation set. Finally, referral formats and documentation habits vary across facilities and change over time, so the rule catalogue and any learned components require periodic review rather than a one-time validation. Privacy, consent, and data-handling requirements must also be addressed before any version of the system processes real patient documents.

XII. FUTURE SCOPE

Future versions could combine document-layout analysis with OCR so that the system better distinguishes tables, forms, and free text, improving field attribution on complex multi-page referrals. Another direction is longitudinal tracking comparing a patient's referrals over time to detect recurring gaps or drifting documentation quality at a given facility, rather than evaluating each referral in isolation. Deeper FHIR/ABDM sandbox integration would allow validated referrals to be exchanged directly with receiving health information systems. Explainable-AI tooling that shows exactly which extracted span drove a given flag would further help clinicians and researchers audit the system rather than rely on it as an unexplained label generator. In the longer term, a shared benchmark dataset of synthetic Indian clinical referrals could let researchers compare extraction and validation approaches under comparable conditions.

XIII. CONCLUSION

Clinical referral and handover communication remains a practical challenge because important information can be missing, contradictory, or ambiguously summarized without this being obvious to the receiving clinician. Structured-communication research has shown that tools such as SBAR help when faithfully implemented, while clinical-NLP research has shown that language models can extract structured content but are not inherently faithful to their source [1]– [5]. This paper proposed ClinBridge, a hybrid framework that combines LLM-based extraction with deterministic completeness and consistency validation, and that converts the result into a source-traceable SBAR draft rather than an unexplained AI summary. The framework should be treated as a documentation-quality screening and evidence-generation tool, not a diagnostic or treatment-decision system. Future implementation and testing should focus on the synthetic scenario set described in Section 7, on OCR-limited real-world documents, and on evaluation with practicing clinicians, so that any eventual claim of clinical usefulness rests on evidence rather than architecture alone.

REFERENCES

[1] S. McCarthy, A. Motala, E. Lawson, and P. G. Shekelle, “Use of structured handoff protocols for within-hospital unit transitions: A systematic review from Making Healthcare Safer IV,” *BMJ Quality & Safety*, 2025.

[2] L. Lo, L. Rotteau, and K. Shojania, “Can SBAR be implemented with high fidelity and does it improve communication between healthcare workers? A systematic review,” BMJ Open, vol. 11, no. 12, Art. no. e055247, 2021.

[3] P. H. Pucher, M. J. Johnston, R. Aggarwal, S. Arora, and A. Darzi, “Effectiveness of interventions to improve patient handover in surgery: A systematic review,” Surgery, vol. 158, no. 1, pp. 85–95, 2015.

[4] M. Agrawal, S. Hegselmann, H. Lang, Y. Kim, and D. Sontag, “Large language models are few-shot clinical information extractors,” in Proc. 2022 Conf. Empirical Methods in Natural Language Processing (EMNLP), Abu Dhabi, UAE, 2022, pp. 1998–2022.

[5] S. Fu, D. Chen, H. He, S. Liu, S. Moon, K. J. Peterson, F. Shen, L. Wang, Y. Wang, and A. Wen, “Clinical concept extraction: A methodology review,” Journal of Biomedical Informatics, vol. 109, Art. no. 103526, 2020.

[6] HL7 International / National Resource Centre for EHR Standards, “FHIR Implementation Guide for ABDM Health Data Interchange Specifications,”

v6.5.0. [Online]. Available: National Resource Centre for EHR Standards

[7] National Health Authority, Government of India, "Ayushman Bharat Digital Mission." [Online]. Available: Ayushman Bharat Digital Mission

[8] National Resource Centre for EHR Standards, "ServiceRequest FHIR Implementation Guide for ABDM v6.5.0." [Online]. Available: ServiceRequest FHIR Implementation Guide