Natural Language Processing for Personalized Healthcare Customer Messaging Using Multilingual Clinical and Service Text

Bilal Ahmed[1], Kamran Hussain[2], Ariful Islam[3]

[1]Department of Environmental Sciences, University of Turbat, Syed Hashmi Avenue, Turbat-92600, Pakistan
[2] Department of Economics, Kohat University of Science and Technology, Bannu Road, Kohat-26000, Pakistan
[3]Dhaka, Bangladesh

ABSTRACT Personalized customer messaging in healthcare increasingly relies on natural language processing to bridge clinical language, service interactions, and patient-facing communication across many languages. Healthcare organizations must communicate appointment logistics, medication reminders, follow-up instructions, billing clarifications, and service recovery messages while adapting tone, reading level, and cultural framing to an individual. At the same time, message generation must be constrained by privacy, safety, and regulatory requirements, and must not blur the boundary between informational support and medical advice. This paper examines technical foundations for building multilingual NLP systems that transform heterogeneous clinical and service text into safe, personalized customer messaging. It analyzes data pipelines that combine structured records with free-text notes and contact-center transcripts, representation learning for cross-lingual clinical semantics, and controlled generation methods that enforce factuality, provenance, and policy constraints. The discussion emphasizes personalization that is clinically and operationally grounded, including preference learning from interaction histories, dynamic user state estimation, and uncertainty-aware decision rules that govern when to ask clarifying questions or route to human staff. Evaluation is treated as a multi-objective problem spanning semantic fidelity, readability, cultural and linguistic adequacy, bias and equity, and safety-related failure modes such as hallucinated instructions or inadvertent disclosure. Deployment considerations include monitoring drift across languages and sites, audit logging, redaction, and governance workflows that encode institutional policy. The goal is to present a cohesive modeling and systems view for multilingual healthcare messaging that is rigorous while remaining compatible with real-world constraints.

I. Introduction

Healthcare customer messaging sits at a complicated interface between clinical content and service operations [1]. A single outbound message may need to incorporate scheduling constraints from the appointment system, clinical constraints from a care plan, administrative constraints from insurance and billing, and individualized preferences such as preferred language, channel, formality, accessibility needs, and reading level. Unlike generic personalization in retail or entertainment, healthcare personalization can carry safety implications because small linguistic shifts can change perceived urgency, implied authorization, or the interpretation of instructions. The technical problem is therefore not simply to generate fluent text, but to generate text that is correct relative to authoritative sources, aligned with institutional policy, consistent with a patient's context, and appropriate within cultural and linguistic norms. Multilingual settings amplify these demands because translation and localization are not only lexical mapping problems; they involve pragmatic choices about politeness, honorifics, indirectness, and culturally specific ways of referencing time, family involvement, and clinical concepts.

The inputs that drive customer messaging are heterogeneous. Clinical text includes clinician notes, discharge summaries, problem lists, radiology impressions, and medication instructions [2]. Service text includes call-center logs, chat transcripts, email threads, complaint narratives, and prior outbound campaigns. These sources differ in style, abbreviations, code-switching patterns, and degrees of formality. Clinical notes can be telegraphic and filled with institution-specific shorthand, while service interactions may contain colloquialisms, emotional expressions, and partial informa-

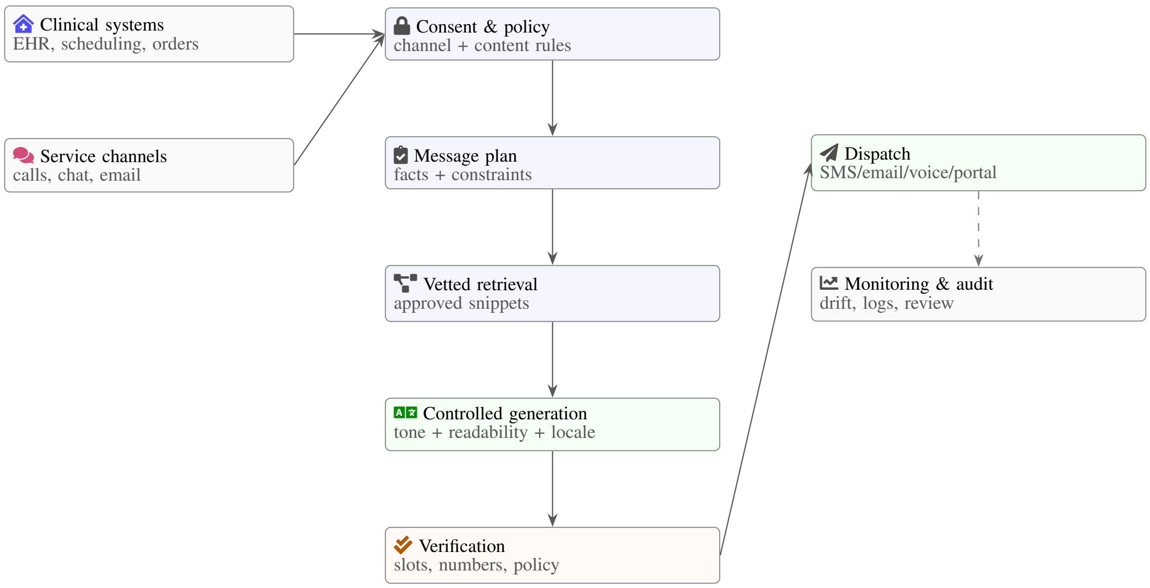

FIGURE 1: End-to-end architecture for multilingual healthcare customer messaging: heterogeneous clinical and service inputs are filtered by consent and institutional policy, converted into a structured message plan, grounded via retrieval of approved phrasing, adapted through controlled multilingual generation, and validated by verification checks before channel dispatch. Continuous monitoring and audit logging close the loop for governance and incident response.

tion. The modeling challenge is to reconcile these sources into a unified, privacy-preserving representation that supports content selection, tone control, and language generation without losing clinically relevant nuance.

Personalized messaging objectives also differ by use case. Operational messages may be constrained to administrative content, such as reminders to bring identification, instructions for fasting prior to a procedure, or clarifications about copays. Engagement messages may aim to improve adherence, encourage follow-up, or support patient navigation [3]. Service recovery messages may need to acknowledge dissatisfaction, apologize without admitting liability, and propose next steps. A single patient may interact across all of these contexts, creating longitudinal histories that influence preference inference and message strategy. For example, if a patient routinely responds to SMS in a particular dialect and prefers concise messages, the system should adapt. If prior interactions reveal confusion about medical terminology, the system should simplify language and offer clarification routes. If the patient has limited digital access or requests paper communication, the channel strategy should change. Personalization therefore spans not only surface realization but also action selection, channel optimization, and escalation decisions [4].

Multilingual clinical and service text introduces distinct difficulties that are not fully captured by general cross-lingual NLP benchmarks. Clinical language is specialized, with domain-specific named entities, negation and temporality patterns, and frequent use of abbreviations. In many languages, clinical terminology is borrowed from English or Latin, but the borrowed forms vary across regions, creating mismatches that standard bilingual dictionaries do not capture. Service text adds pragmatic variability, including slang, code-switching, and speech-to-text artifacts. Many healthcare organizations also operate in regions where low-resource languages are common, and where orthographic norms vary or are not standardized. The system must handle mixed scripts, informal romanization, and evolving terminology. These factors make it risky to rely solely on generic multilingual language models without domain adaptation and safety scaffolding [5].

The notion of personalization itself requires careful definition. A naive approach that conditions generation directly on a patient profile can inadvertently leak sensitive attributes, infer conditions, or create messages that feel intrusive. Effective personalization in healthcare must be constrained to attributes that are appropriate for communication and grounded in consent. It must also respect fairness considerations, because personalization based on historical data can reinforce inequities in access, engagement, or clinical outcomes. For instance, if historically underserved communities have lower engagement due to systemic barriers, a model that predicts low engagement could allocate fewer resources to those communities, creating feedback loops. Personalization should instead be designed as a policy that aims for equitable service quality while adapting to individual constraints [6].

This paper focuses on the technical design of NLP systems for personalized healthcare customer messaging using multi-

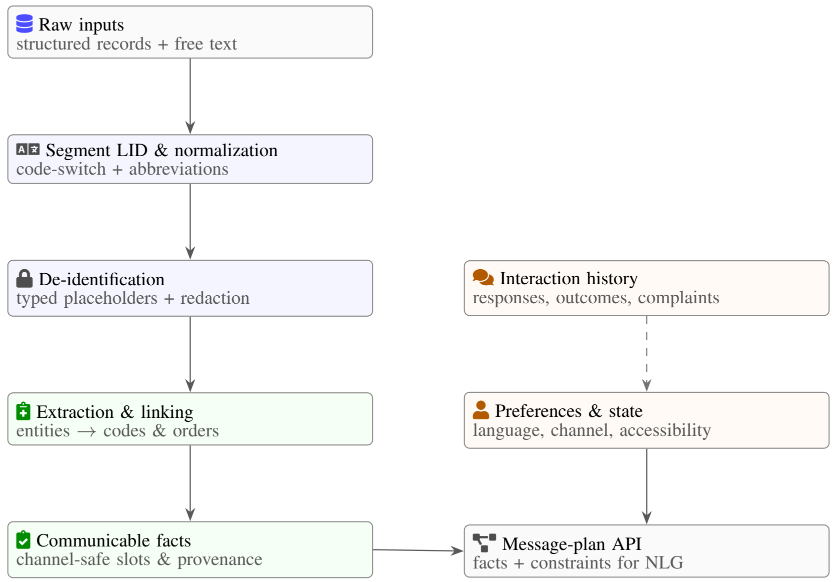

FIGURE 2: Data and representation pipeline that separates raw clinical/service text from communication-appropriate content. Fine-grained language identification, normalization, and privacy-preserving de-identification feed extraction and entity linking to produce a curated fact set with provenance, joined with explicit preference/state features to construct message plans for downstream retrieval and controlled multilingual realization.

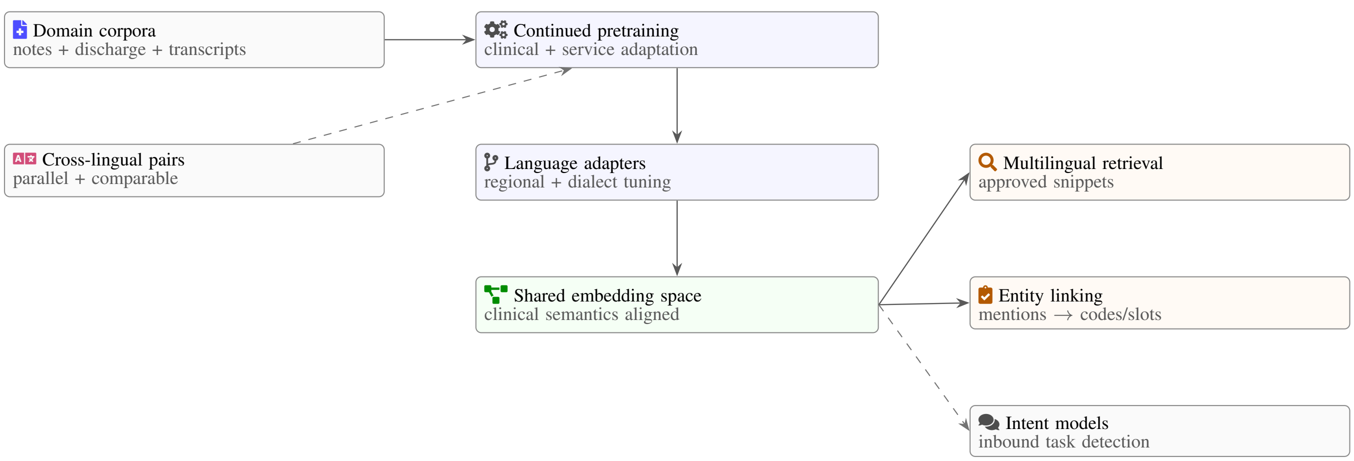

FIGURE 3: Cross-lingual representation strategy for clinical and service language: domain-adaptive pretraining and lightweight language/dialect adapters produce aligned embeddings for specialized terminology. The shared space supports multilingual retrieval of vetted phrasing, robust entity linking into structured codes/slots, and calibrated intent models for inbound requests.

lingual clinical and service text. The framing is end-to-end, covering data curation and privacy, representation learning across languages and domains, controlled generation and retrieval, preference and state modeling, evaluation methodologies, and deployment governance. While the term customer messaging may evoke marketing, the scope here centers on informational, navigational, and service communications that support care delivery, operations, and patient experience. The discussion assumes that the system operates within an organizational environment with clear policy boundaries, human oversight pathways, and legal constraints. Within that environment, the technical goal is to reduce friction and improve clarity without overstepping into unverified medical guidance.

A key architectural question is whether messaging should be generated from scratch, templated, retrieved, or hybrid. Pure templating offers predictability but limited personalization and multilingual flexibility [7]. Retrieval-based approaches can reuse vetted messages but may fail to fit context or language variety. Generative models offer flexibility but introduce risks of hallucination, unintended implications, or inappropriate tone. Hybrid systems that retrieve approved content and then perform constrained adaptation are increasingly practical because they combine stability with

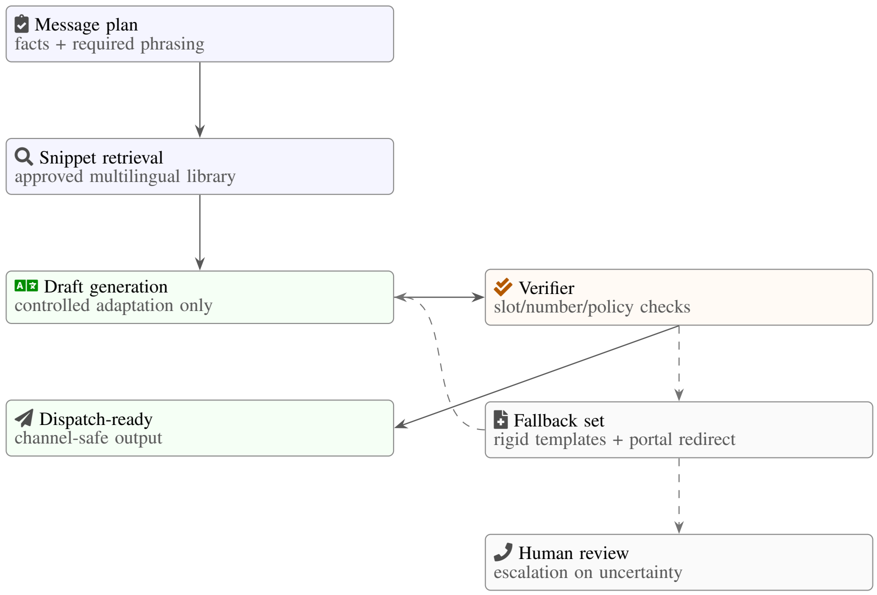

FIGURE 4: Hybrid retrieval-and-generation workflow with explicit verification. A structured message plan and approved snippets ground a controlled draft; a verifier enforces slot consistency, numeric/temporal correctness, and channel/policy constraints. Failed checks route to conservative fallbacks (templated redirects) or to human review, limiting the risk of hallucinated or inappropriate instructions across languages.

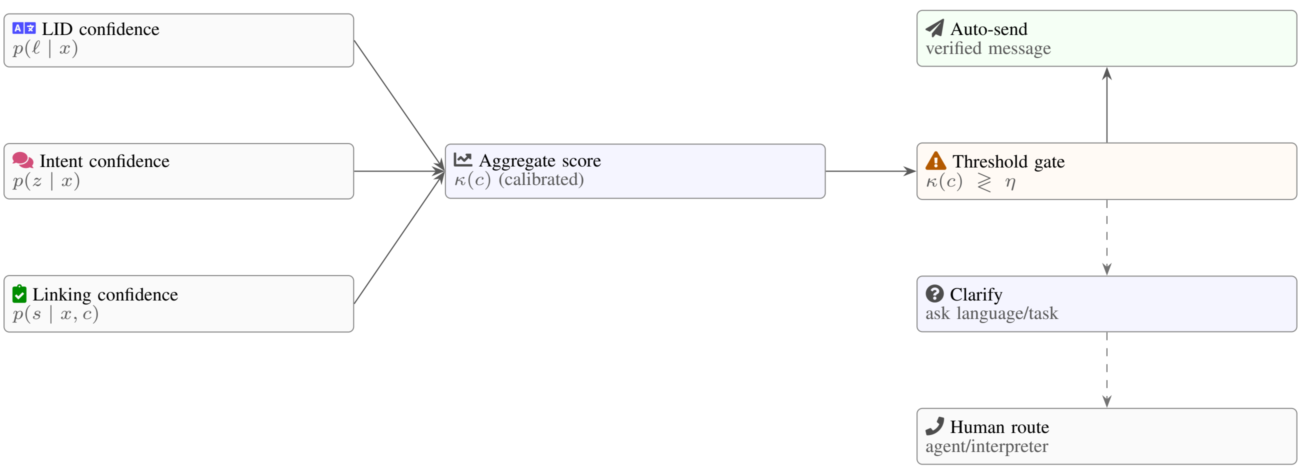

FIGURE 5: Uncertainty-aware routing for multilingual customer messaging. Calibrated confidence signals from language identification, intent detection, and entity/slot linking are aggregated into a single risk score that gates automation. Low-confidence cases trigger clarification prompts or escalation to human staff, reducing unsafe or inequitable failures when language resources are uneven.

contextual tailoring. Hybridization also supports multilingual workflows, where human-translated canonical messages exist for some languages, and models adapt phrasing to patient preferences while maintaining meaning.

Another central question is how to integrate clinical provenance. A message that references a medication dose must trace to an authoritative medication order, not a speculative inference from a note. A reminder about follow-up timing must reference the scheduling system and the clinician's plan [8]. The model must therefore be embedded in a broader system that exposes structured facts and links them to message spans. This motivates retrieval-augmented generation over curated knowledge stores, as well as span-level attribution that can be audited. Multilinguality complicates attribution because the generated language differs from the source language, so the system must maintain semantic alignment between source facts and target phrasing.

The remainder of the paper develops these points. It first articulates the problem setting and constraints specific to multilingual healthcare messaging, emphasizing safety, privacy, and policy compliance. It then discusses data sources and representation strategies that unify clinical and service

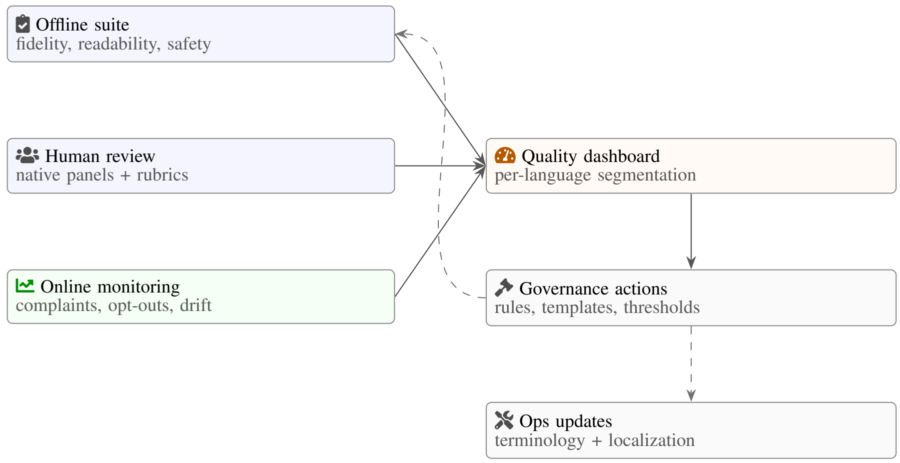

FIGURE 6: Evaluation and governance loop for multilingual healthcare messaging. Offline test suites and structured human review feed a per-language dashboard alongside online monitoring signals (delivery outcomes, complaint categories, and drift indicators). Governance workflows translate these measurements into controlled updates to policies, templates, thresholds, and localization resources.

Dimension	**Definition in this setting**	**Illustrative examples**
Context	Structured and unstructured information about the person and episode of care used to drive messaging decisions.	Demographics and language preference; upcoming appointments; open tasks; interaction history across SMS, portal, phone.
Operational objective	Explicit task the organization is trying to achieve with a message or action.	Reduce no-show rates; clarify pre-procedure requirements; resolve billing questions; support navigation after discharge.
Action space	System-level choice that may include or bypass message generation.	Send reminder; ask clarifying question; schedule call-back; escalate to human agent; direct to portal.
Message realization	Linguistic form that encodes selected facts, tone, and channel constraints.	Short SMS vs detailed email; formal vs warm-professional; monolingual vs bilingual messages.

TABLE 1: Core elements of the healthcare customer messaging problem formulation.

text while controlling risk [9]. Next it details modeling approaches for multilingual understanding, personalization, and controlled generation, including light-weight mathematical formulations for multi-objective optimization and uncertainty-aware routing. It then examines evaluation methods and failure analysis tailored to healthcare messaging, focusing on semantic fidelity, safety, and equity across languages. Finally it addresses deployment, monitoring, and governance, emphasizing the practical mechanisms that make such systems reliable in production.

II. Background and Problem Setting

Healthcare customer messaging can be formalized as a decision-and-generation process that maps a user context and an operational objective to a message and an action. The context includes demographic and preference information that the patient has consented to share for communication purposes, the current episode of care, operational status such as upcoming appointments and outstanding tasks, and the interaction history across channels. The objective can be operational, such as reducing no-show rates, clarifying pre-procedure requirements, or resolving billing questions. The

Source type	Typical content	Linguistic properties	Messaging relevance
Clinical notes	Assessments, plans, diagnostic reasoning, differential diagnoses, social history.	Telegraphic style, dense shorthand, institution-specific abbreviations, domain-specific negation and temporality.	Provide clinical constraints on what can be safely referenced (e.g., finalized plan vs speculation).
Discharge and instruction fields	Standardized instructions, problem lists, preparation steps, follow-up recommendations.	Semi-structured templates mixed with free text; partial localization across languages.	Primary source for actionable instructions and follow-up logistics.
Service transcripts (call/chat)	Questions, complaints, negotiations about scheduling, billing, access barriers.	Colloquialisms, emotion, code-switching, ASR artifacts, incomplete turns, interruptions.	Reveal confusion, preferred tone, and pain points that shape personalization strategy.
Outbound campaigns and prior messages	Reminders, engagement nudges, service recovery communications.	Mixture of hand-authored, templated, and manually translated text with uneven quality.	Historical supervision for content planning, tone, and multilingual realization.

TABLE 2: Heterogeneous clinical and service text sources that feed personalized messaging.

Level	Multilingual challenge	Consequences for messaging	Mitigation strategies
Lexical and terminological	Borrowed clinical terms with regional variants; low-resource vocabularies; inconsistent orthography.	Mismatched terminology; misunderstanding of procedures or medications; uneven clarity across languages.	Domain adaptation; multilingual lexicons; human-curated canonical phrases per language.
Pragmatic and cultural	Different norms for politeness, honorifics, directness, and family involvement.	Messages may sound rude, overly intimate, or confusing even when literally correct.	Language-specific tone guidelines; culturally informed templates; native-speaker review.
Code-switching and mixed scripts	Frequent mixing of languages, romanization of non-Latin scripts, embedded English medical terms.	Language identification errors; degraded intent classification and extraction; brittle rules.	Fine-grained language ID; robust tokenization; training on real mixed-language corpora.
Quality and equity	High-quality human translations for some languages; reliance on generic models for others.	Systematic differences in comprehension and safety; increased routing to humans for certain groups.	Per-language quality estimation; explicit fallbacks; investment guided by equity monitoring.

TABLE 3: Key multilingual challenges beyond generic cross-lingual NLP benchmarks.

action may be to send a message, ask a clarification question, schedule a call-back, or escalate to a human agent [10]. The message itself is a linguistic realization that must satisfy constraints on content, tone, and policy.

This setting has structural asymmetries that are important for modeling. First, the organization often has more data about the patient than the patient wants surfaced in a message. A message that reveals sensitive diagnoses, test results, or family information can violate privacy expectations even if the organization legitimately holds the data. Second, the patient's interpretation is influenced by trust and anxiety, meaning that ambiguity can cause harm even when content is technically correct. Third, the organization bears responsibility for accuracy and compliance, so conservative behavior may be preferable, such as routing uncertain cases to humans [11]. These asymmetries imply that the system objective should not be framed as maximizing personalization intensity, but as maximizing clarity and task completion subject to safety and privacy constraints.

Layer	Primary role	Multilingual considerations	Safety and privacy hooks
Language and segmentation preprocessing	Detect language(s), segment notes and transcripts, normalize punctuation and casing.	Fine-grained language tags for code-switched segments; support for mixed scripts and transliteration.	Avoid storing raw identifiers; early redaction of unsafe free-form fields.
Extraction and de-identification	Map raw text to communicable facts; remove or mask identifiers.	Multilingual NER and entity linking; placeholders that respect morphology and script.	Typed placeholders for names and contacts; separation of speculative vs confirmed facts.
Cross-lingual representation	Embed entities, segments, and interaction histories into shared semantic spaces.	Alignment of clinical entities across languages; handling colloquial medication descriptions.	Constrain mappings to patient-specific candidate sets; explicit uncertainty tracking.
Message plan construction	Select which facts and options can be surfaced given context and objective.	Language-independent plan representation; later localized per language and channel.	Enforce consent and policy filters; exclude sensitive or restricted content by design.

TABLE 4: Data and representation pipeline layers supporting multilingual messaging.

Personalization axis	Driving signals	Intended adaptation	Associated risks
Language and dialect	Recorded language preference; observed reply language; regional metadata.	Choice of language, dialectal vocabulary, and politeness conventions.	Lock-in to incorrect preference; inconsistent quality across languages.
Tone and formality	Historical responses; channel; age band; institution branding guidelines.	Neutral vs warm-professional tone; level of empathy; degree of directness.	Over-familiar or insincere phrasing; unintended promises or implications.
Readability and detail level	Patterns of confusion; explicit requests for simpler language; channel constraints.	Sentence length, use of jargon, inclusion of optional explanations or links.	Oversimplification of clinically important nuance; inferred literacy proxies.
Channel and timing	Device access; prior engagement; consented channels; operational objectives.	SMS vs email vs call; proactive vs reactive outreach; batching of reminders.	Channel leakage of sensitive content; message fatigue; inequitable outreach.

TABLE 5: Major personalization dimensions and their communication-oriented trade-offs.

Multilinguality changes the operational definition of correctness. In monolingual settings, correctness can be measured relative to source facts and language norms. In multilingual settings, correctness includes equivalence of meaning across languages and dialects, appropriateness of register, and cultural pragmatics. For example, some languages require explicit politeness markers in service contexts, while others rely on indirectness to avoid appearing confrontational. Direct translations that preserve propositional content can still be pragmatically wrong, leading to dissatisfaction or misunderstanding. This implies that evaluation must incorporate pragmatics and interaction outcomes, not only semantic similarity [12].

Clinical language also interacts with multilinguality in complex ways. Many clinical concepts lack commonly understood equivalents in some languages, especially when patient-facing language differs from clinician-facing terminology. Patients may recognize borrowed technical terms but not their implications, or they may prefer descriptive phrases. Additionally, certain concepts are taboo or culturally sensitive, requiring careful phrasing. For example, mental health or reproductive topics may require more discretion in

Component	Function	Design emphasis	Examples in setting
Intent and task detection	Map inbound free text to operational intents and routes.	Multilingual encoders, calibration, abstention and human escalation.	Detect rescheduling vs cancellation; distinguish billing questions from clinical queries.
Content planner	Assemble structured message plans from authoritative systems.	Determinism for critical fields; explicit consent and policy rules.	Combine appointment time, location, preparation, and contact options into a plan.
Retrieval of vetted phrasing	Select library snippets consistent with plan, language, and tone.	Multilingual semantic search; template tagging by context and register.	Reuse pre-approved fasting instructions or apology patterns across channels.
Controlled generator and verifier	Adapt snippets to context while enforcing meaning preservation.	Constrained paraphrase; lexical constraints; cross-lingual factuality checks.	Insert patient-specific dates; adjust formality; verify entity consistency before send.

TABLE 6: High-level modeling components in a hybrid decision-and-generation architecture.

Constraint type	Failure modes	Preventive mechanisms	Fallback behaviors
Privacy and consent	Revealing sensitive diagnoses; including restricted details on insecure channels.	Upstream content filters; channel-aware policies; minimal necessary context retrieval.	Send generic portal notification; omit details; route to authenticated channel.
Clinical safety and factuality	Hallucinated instructions; incorrect dosage, timing, or authorization.	Grounding in structured orders; numeric and temporal consistency checks; entailment verifiers.	Revert to rigid template; human review for low-confidence cases; abstain from clinical commentary.
Policy and regulatory boundaries	Crossing from informational messaging into medical advice or triage.	Intent classifiers; content filters for prescriptive language; explicit exclusion of diagnostic reasoning.	Acknowledge questions; direct to clinicians or nurse line; log for follow-up.
Equity across languages	Lower quality or more frequent failures for low-resource languages or dialects.	Per-language metrics; routing thresholds tuned with fairness constraints.	Prioritize interpreter support; adjust automation coverage; targeted model improvement.

TABLE 7: Safety, privacy, and equity constraints and how they shape system behavior.

some contexts. A messaging system must therefore support controlled paraphrasing and optionality, such as providing a general reminder to check the patient portal rather than naming a sensitive test [13]. These decisions should be governed by policy and consent rather than model improvisation.

Service text introduces conversational dynamics that differ from clinical documentation. Call-center transcripts and chat logs contain incomplete sentences, emotional expressions, and references to earlier turns. They also contain negotiation and conflict, such as disputes about charges or frustration about wait times. Personalization in these contexts includes empathy and de-escalation phrasing, but it must be carefully bounded to avoid making promises the organization cannot keep. In multilingual settings, expressions of empathy vary widely; literal translations may sound insincere or overly familiar. Therefore, the model must be trained or guided with domain-specific conversational data and institutional tone guidelines [14].

A practical way to conceptualize personalized messaging is as constrained natural language generation conditioned on a structured set of facts and decisions. The system first determines what facts to include, such as appointment time and location, required preparation steps, contact numbers, or next actions. It then chooses a communication strategy, including tone, level of detail, and whether to provide optional clarifications. It then realizes the message in the target language and channel format, such as SMS character limits or voice call scripts. At each stage, constraints apply. Fact

Evaluation axis	Key questions	Methods	Multilingual considerations
Semantic fidelity and factuality	Does the message match the plan and authoritative records?	Slot extraction and comparison; numeric/date checks; cross-lingual entailment models.	Language-specific date and numeral formats; differing negation and modality markers.
Readability and comprehension	Can typical recipients understand and act on the message?	Readability proxies; human comprehension tests; task completion studies.	Language-specific formulas; lack of tools for some languages; culturally grounded examples.
Tone, pragmatics, and satisfaction	Is the message polite, respectful, and aligned with expectations?	Native-speaker ratings; complaint analysis; A/B tests under ethical constraints.	Honorifics and politeness systems; indirectness preferences; family-oriented phrasing.
Safety, privacy, and equity	Are violations rare and equitably distributed across groups?	Adversarial testing; audit sampling; per-language error and escalation analysis.	Different routing patterns by language; interpreter availability; heterogeneous training data.

TABLE 8: Evaluation dimensions for personalized multilingual healthcare messaging.

Deployment area	Operational focus	Concrete mechanisms	Multilingual-specific issues
Identity and consent management	Ensure messages reach the right person with appropriate content.	Verified contact points; role-based consent; per-channel and per-topic flags.	Caregiver vs patient language; consent capture in multiple languages; mixed-language households.
Monitoring and drift detection	Track system health and detect degradation or attacks.	Verification failure dashboards; canary releases; anomaly detection on outputs.	Per-language KPIs; new dialects at specific sites; prompt injection expressed in different languages.
Governance and oversight	Define automation scope, review processes, and escalation rules.	Message category whitelists; model versioning; structured human review queues.	Uneven interpreter capacity; language onboarding checklists; equity audits by language.
Localization and terminology operations	Maintain consistent naming, phrasing, and references across languages.	Multilingual terminology databases; approved phrasing libraries; update workflows.	Region-specific clinic names; local transport and time expressions; evolving medical vocabulary.

TABLE 9: Deployment, monitoring, and governance considerations for real-world rollout.

selection is constrained by relevance and privacy. Strategy is constrained by institutional policy, equity goals, and user preferences [15]. Realization is constrained by language norms and safety phrasing.

The consent and privacy boundary is central. Patients may consent to receiving messages via certain channels and to receiving certain categories of information. For example, some patients may allow appointment reminders but not medication details via SMS. Some may permit language personalization but not the use of inferred attributes. In multilingual contexts, consent itself must be communicated and recorded across languages [16]. From a systems standpoint, consent should be represented as structured, machine-enforceable policies, and the generation component should receive only the permitted facts. This reduces the risk that a model will reveal restricted information because it never sees it.

Equity considerations arise when multilingual support is uneven. Healthcare organizations often have high-quality human translations for a subset of languages and rely on automated methods for others. This can lead to differential message quality, affecting comprehension and outcomes. A system should therefore include explicit quality estimation and fallback mechanisms. If the model's confidence is low for a given language or dialect, it may be safer to route to a human interpreter or to send a minimal message directing the patient to a staffed line [17]. This introduces a trade-off between access and precision that should be managed explicitly rather than implicitly through model failure.

Another important constraint is interpretability and auditability. Healthcare communication often requires the ability to explain why a message was sent and what information it was based on. This is not only a compliance need but also operationally useful for handling complaints and improving workflows. In multilingual messaging, auditability includes being able to show how a message in one language corresponds to source records potentially in another language. This motivates maintaining structured message plans, span-level provenance links, and versioning of templates or model prompts [18].

The problem setting also includes dynamic context updates. Appointment times change, medication orders are modified, and insurance authorizations are updated. A messaging system must be robust to stale data and must avoid sending messages based on outdated records. This implies tight integration with operational systems and careful caching policies. In multilingual environments, latency and synchronization issues can be more pronounced when translation or generation services add delay. Therefore, system design must prioritize real-time access to authoritative sources and incorporate validation checks immediately before dispatch.

Finally, personalized messaging must respect boundaries between informational communication and medical advice [19]. A model that elaborates on symptoms or provides clinical recommendations can cross into regulated clinical decision support. Even when the underlying data are present, a message that interprets results or suggests treatment changes can be inappropriate. The system should therefore implement content filters and intent classifiers that prevent diagnostic or prescriptive content unless explicitly authorized and verified. When patients ask clinical questions in inbound channels, the system can acknowledge and route to clinicians rather than generating medical guidance.

III. Data and Representation for Multilingual Clinical and Service Text

Data for personalized healthcare messaging typically spans structured records, semi-structured fields, and unstructured text. Structured data includes appointments, provider identifiers, clinic locations, medication orders, procedure codes, and billing statuses [20]. Semi-structured data includes problem lists, discharge instruction fields, and templated note sections. Unstructured data includes clinician narrative notes, patient messages, call transcripts, and chat logs. In multilingual contexts, any of these can appear in multiple languages, and some records may contain code-switching within a single document. Data preparation must therefore address language identification at fine granularity, normalization of abbreviations, and alignment of entities across languages.

A core design principle is separating private clinical detail from communication-appropriate facts. Clinical notes may contain speculative diagnoses, differential considerations, or sensitive social context. Service text may contain personal anecdotes or third-party information [21]. Not all of this is appropriate to surface. Therefore, an extraction layer should map raw text to a curated set of communicable facts. This layer can combine rule-based extraction for high-precision fields with machine learning extraction for broader coverage. In multilingual settings, extraction can be performed either in the source language with multilingual models, or by translating to a pivot language and extracting there. Each approach has trade-offs. Direct multilingual extraction preserves nuance but requires robust models across languages. Pivot translation can leverage stronger monolingual extractors but introduces translation errors and can distort entities such as medication names or addresses [22]. A practical compromise is to perform entity recognition and key field extraction in the source language while translating only for downstream semantic tasks that tolerate noise.

De-identification and privacy-preserving representation are essential. Service transcripts and clinical notes contain identifiers, addresses, phone numbers, and unique combinations of facts that can re-identify individuals. For training and evaluation, organizations often create de-identified corpora by removing direct identifiers and masking quasi-identifiers. However, de-identification can degrade language modeling performance if it removes clinically meaningful tokens. A better approach is to replace identifiers with typed placeholders that preserve syntactic behavior, such as replacing a person name with a token indicating a name, and replacing a phone number with a phone placeholder [23]. In multilingual contexts, placeholders should be language-agnostic, and the surrounding morphology must be handled, for example with clitic attachment or case marking. If a language attaches suffixes to names, placeholders must support that morphology, potentially by using subword representations and allowing the model to generate appropriate suffixes around the placeholder.

Another representation challenge is aligning clinical entities across languages. Medication names, procedure names, and diagnostic terms may have multiple aliases and spelling variants. Cross-lingual lexicons can help, but in practice many variants appear in service text, including misspellings and colloquial descriptions. Representation learning can mitigate this by embedding terms in a shared semantic space across languages. However, embeddings alone are not sufficient for safety-critical messaging because ambiguity must be resolved [24]. For example, a colloquial reference to a medication class must be mapped to the specific order in the record before it is mentioned in a message. This motivates entity linking to structured codes when possible. For multilingual text, entity linking can be performed via multilingual encoders that map mentions to candidate codes using contextual similarity, with constrained candidate sets derived from the patient's active orders to reduce ambiguity.

Clinical and service text also differ in discourse structure. Clinician notes may contain headings, assessment-plan structure, and temporal references. Service transcripts contain turn-taking and may include interruptions or speech recognition artifacts [25]. A unified representation should preserve these structures because they influence what is safe to infer. For example, a clinician note may contain a plan to discuss a topic later, which should not be interpreted as a finalized instruction. A transcript may contain a patient's misunderstanding that should not be repeated. Therefore, representations should encode source type, section boundaries, and speaker roles. For multilingual transcripts, speaker role identification can be complicated when bilingual staff switch languages mid-call. Fine-grained language tags per segment can support more accurate modeling.

From a training perspective, one must decide which parts of the pipeline are learned and which are rule-governed [26]. For safety and compliance, it is often preferable for the content selection layer to be deterministic for certain fields, such as appointment time and location, because errors are costly. Learned components can then focus on paraphrasing, tone adaptation, and multilingual realization. This division reduces the burden on the language model and provides clearer audit trails. It also supports incremental improvement, because the deterministic layer can be updated with new policy rules without retraining the model.

Building a multilingual dataset for messaging requires pairing contexts with target messages. Historical outbound messages provide natural supervision, but they may reflect legacy quality issues, inconsistent tone, and unequal translation quality across languages [27]. Some messages may have been authored by staff, others by templating systems, and others by ad hoc manual translation. Training directly on this data can perpetuate inequities and errors. Data curation should therefore include quality stratification, where messages are scored for clarity, correctness, and policy compliance. In multilingual settings, quality scoring can incorporate human review for a subset of languages and automatic heuristics for others, such as checking consistency of extracted facts with the message. Quality-aware sampling during training can reduce the influence of low-quality examples.

Synthetic data can also be useful, but it must be carefully designed. One approach is to generate message plans from structured data and then have professional translators produce messages in target languages for a representative set of plans [28]. Another approach is to create controlled paraphrases in each language to teach tone variation and readability adjustment. Synthetic augmentation can also include noise injection that simulates speech recognition errors or misspellings, improving robustness when parsing service text. However, synthetic generation should avoid creating clinically implausible combinations, and should be constrained to policy-safe content. In practice, synthetic data is most effective when it is anchored to real operational distributions, such as typical appointment types, typical preparation instructions, and common service issues.

A representation that supports personalization must also encode user preferences and state. Preferences include language, dialect, channel, formality, and accessibility requirements such as large-print or simplified language. State includes recent events such as missed appointments, recent complaints, or ongoing authorization issues [29]. These variables should be represented explicitly rather than inferred implicitly from text, because explicit representation allows consent control and easier auditing. For multilingual messaging, the language preference should include both language and regional variant when possible, because vocabulary and politeness norms differ. The system should also support uncertainty in preferences, such as when a patient's language preference is unknown or when the patient uses multiple languages. In such cases, the system can adopt a conservative strategy, such as sending bilingual messages or asking for preference confirmation, while respecting channel constraints.

Privacy-preserving learning techniques can be relevant when training on sensitive text. Approaches such as differential privacy, secure enclaves, and federated learning can reduce exposure of raw data [30]. In practice, the choice depends on organizational constraints and the feasibility of engineering such systems. Even without advanced privacy learning, strong controls can be applied through access governance, data minimization, and rigorous de-identification. For multilingual data, privacy risk can vary across languages if some languages encode personal information differently, such as including honorifics that imply gender or social status. De-identification and policy rules should therefore be evaluated per language.

Finally, multilingual customer messaging often requires handling code-switching and mixed scripts. Patients may write in one language using another script, such as romanized forms of non-Latin languages. They may also mix English medical terms into another language [31]. Models trained only on standard corpora may struggle with these patterns. Data collection should therefore include real-world mixed-language examples, and tokenization should support mixed scripts robustly. Subword tokenization can help, but it can also fragment rare scripts excessively, reducing performance. In such cases, character-level components or byte-level encoding can improve robustness, though they may increase computational cost. The representation strategy should be chosen based on observed language patterns in the service population rather than theoretical coverage.

IV. Modeling Approaches for Personalized Multilingual Messaging

A production-grade system for personalized multilingual healthcare messaging is typically composed of multiple models and rule layers rather than a single end-to-end generator [32]. A useful decomposition includes intent and task detection for inbound messages, content planning from structured sources, retrieval of vetted phrasing, controlled generation for adaptation, and safety verification before dispatch. Each component can be multilingual, but not all must be equally capable across languages if the system includes fallbacks and routing. The design goal is to allocate learning capacity to the parts where flexibility matters, while constraining the parts where errors are costly.

For inbound messaging, the system must map user text to intents such as appointment rescheduling, billing questions, medication refill requests, or complaints. Multilingual intent classification can be built using multilingual encoders fine-tuned on labeled service data. In low-resource languages, cross-lingual transfer can be effective if the intent space is shared and if training includes code-switched examples. Intent models should be calibrated, because uncertain classifications should route to human agents rather than triggering automated responses [33]. In practice, calibration can be improved through temperature scaling or isotonic regression on a held-out set per language. Because language distributions differ across sites, calibration should be monitored for drift.

Content planning for outbound messages should prioritize deterministic extraction from authoritative systems. For example, appointment reminder content can be assembled from scheduling records, including time, location, provider, and preparation requirements. Clinical preparation requirements can be linked to procedure types and pre-approved instruction sets. The planner can also incorporate patient-specific constraints, such as accessibility needs or transportation arrangements, if those are explicitly recorded and permitted for messaging [34]. The output of the planner is a structured message plan that enumerates facts and allowable phrasing options, along with constraints such as maximum length for SMS or required disclaimers.

Retrieval of vetted phrasing can be implemented using a multilingual semantic search over an approved library of message templates and snippets. Each snippet can be tagged with applicable contexts, languages, tone, and policy constraints. Multilingual retrieval can use dual-encoder embeddings trained on parallel or comparable corpora, with domain adaptation on clinical and service text. Retrieval provides stable phrasing for sensitive content, such as instructions about fasting, arrival times, or billing disputes. It also supports consistent branding and reduces variability that can confuse patients.

Controlled generation then adapts retrieved content to the specific context [35]. Adaptation includes inserting patient-specific facts, adjusting tone and reading level, and applying dialect preferences. The generator can be a multilingual sequence-to-sequence model or a large language model configured with constrained decoding. However, in healthcare messaging, unconstrained generation is risky. A safer approach is to condition generation on the message plan and retrieved snippets, and to restrict output to transformations that preserve meaning. This can be achieved through editing models that perform paraphrase with constraints, or through generation with lexical constraints that ensure critical entities are preserved. Another approach is to use a small set of allowed paraphrase operators, such as simplifying sentences, reordering for clarity, and adding polite greetings, and then applying these operators through learned models or rules.

Personalization requires modeling user preferences and interaction history [36]. One method is to treat personalization as a conditional generation problem where the model receives a preference vector and a user state summary. The preference vector can include language, formality, verbosity, and channel. The user state summary can include recent interactions and outcomes, such as whether the user responded, whether the user clicked a link, or whether the user required human follow-up. To avoid privacy risks, the state summary should exclude sensitive clinical content and should focus on communication behavior. In multilingual contexts, the state summary can include language-switch behavior, such as whether the user replies in a different language than the outgoing message, indicating a mismatch.

Preference learning can be done through supervised signals, such as explicit preferences provided by the user, and implicit signals, such as response rates [37]. However, implicit signals can be confounded by access barriers. Therefore, a careful design frames preference inference as uncertainty estimation rather than hard classification. The system can maintain probability distributions over preferences and update them with Bayesian or quasi-Bayesian methods. For example, if a user replies consistently in one language, the probability of that language preference increases, but the system should still allow easy correction and should avoid locking in a preference based on limited evidence.

The decision of what message to send, when to send it, and whether to include optional details can be framed as a policy optimization problem. The policy chooses an action based on context, and the action yields outcomes such as task completion, reduced no-shows, or reduced call volume. In healthcare, the reward function must incorporate safety and equity constraints [38]. A minimal formalization can treat the policy as maximizing expected utility under constraints. A compact mathematical statement can clarify the multi-objective nature without requiring long equations:

$$\begin{aligned} \max_{\pi} \;& \mathbb{E}\big[U(y,a,c)\big] \\ \text{s.t.} \quad & \mathbb{P}\big(\text{privacyviolation}(y,c)=1\big) \le \epsilon \\ & \mathbb{P}\big(\text{safetyviolation}(y,c)=1\big) \le \delta \\ & \Delta_{\text{lang}}\big(Q(y \mid \ell)\big) \le \tau \end{aligned} \tag{1}$$

Here, π is a policy that selects a message y and action a given context c, while U aggregates operational utility such as clarity and task success. The constraint $\Delta_{\text{lang}}(Q(y \mid \ell))$ represents bounded quality disparity across languages ℓ, operationalized through evaluation metrics. The thresholds ϵ, δ, and τ encode organizational risk tolerance. The point is not to solve this optimization exactly, but to implement approximations through routing, filtering, and per-language quality estimation.

A critical component is factuality control [39]. Messaging systems should avoid hallucinating facts not present in authoritative sources. Retrieval-augmented generation can help by grounding generation on retrieved snippets and structured facts. Yet even with retrieval, models can rephrase incorrectly, especially when translating across languages where tense, aspect, or negation differs. A practical strategy is to generate a draft and then run a verifier that checks the draft against the message plan. Verification can include rule-based checks for entity consistency, numeric consistency, and presence of required disclaimers, as well as learned entailment models that test whether the draft is supported by the plan. For multilingual drafts, entailment models should be cross-lingual or should operate on language-agnostic representations. When verifiers are uncertain, the system can revert to a more rigid template [40].

Readability control is another important modeling aspect. Patient-facing language should match the user's comprehension, which may correlate with education, but using such proxies is ethically fraught. A safer approach is to offer generally clear language, avoid jargon, and optionally provide more details via links or portal content. The model can be trained to simplify text while preserving meaning, using parallel corpora of clinical-to-plain-language transformations. In multilingual settings, simplification resources are uneven, so one can train a multilingual simplifier with shared parameters and language-specific adapters. The simplifier should be constrained to avoid changing numeric values, medication names, or dates [41].

Tone and empathy adaptation can improve experience in service recovery contexts. However, empathy phrasing must be culturally and institutionally appropriate. A model can be conditioned on a tone variable, but the allowed tone space should be bounded, for example ranging from neutral-professional to warm-professional, avoiding overly intimate language. In multilingual settings, tone variables should map to language-specific realizations. This can be achieved by training on annotated examples of tone in each language or by using a small set of vetted tone templates that the generator adapts.

Handling sensitive topics requires policy-aware generation. If the message plan contains sensitive content, such as test results, the system may be required to direct the patient to a secure portal rather than including details in an SMS [42]. Policy can be implemented as a classifier that labels message plans with allowed channels and detail levels, or as a rules engine that maps content types to channel constraints. The generator then receives only permitted content. This reduces reliance on post-hoc filtering, which can fail if the model has already produced sensitive text.

Uncertainty-aware routing is essential. In multilingual systems, uncertainty arises from language identification, intent classification, entity linking, and generation fidelity. Instead of forcing a single model to handle all cases, the system can route based on confidence. For instance, if language identification confidence is low, the system can send a bilingual prompt asking the user to confirm language preference [43]. If intent classification is uncertain, the system can ask a clarification question or route to a human agent. If generation verification fails, the system can revert to a template. This routing logic can be formalized as a threshold policy over calibrated confidence scores. A compact expression can illustrate the idea without long derivations:

$$\begin{aligned} a &= \text{route}(c) \\ \text{route}(c) &= \begin{cases} \text{human} & \text{if } \kappa(c) < \eta \\ \text{auto} & \text{if } \kappa(c) \ge \eta \end{cases} \end{aligned} \tag{2}$$

Here, $\kappa(c)$ denotes an aggregate confidence score derived from upstream components, and η is a risk-tuned threshold. In practice, routing can be multi-level, but the binary form captures the principle that automation should be conditional on confidence [44].

Multilingual modeling choices also include whether to use a single shared model or language-specific models. Shared models benefit from transfer and reduce maintenance overhead, but they may underperform for low-resource languages and can encode biases from dominant languages. Language-specific models can achieve higher accuracy but increase operational complexity. A hybrid approach uses a shared multilingual backbone with language adapters or lightweight fine-tuning per language cluster. This allows targeted improvements without duplicating the full model. For dialect variation, one can use region adapters or lexicon augmentation to capture local vocabulary.

Domain adaptation is necessary because general multilingual models often lack coverage of clinical abbreviations and service jargon [45]. Adaptation can be done via continued pretraining on de-identified clinical and service text, followed by supervised fine-tuning for messaging tasks. However, continued pretraining can increase memorization risk if not carefully managed. Data minimization, placeholder replacement, and privacy-aware training protocols can reduce this risk. Another approach is to keep the base model fixed and learn smaller task-specific layers, which can reduce overfitting to sensitive content.

Finally, controlled multilingual generation should consider channel constraints. SMS requires brevity and careful punctuation because line breaks can render inconsistently across devices [46]. Email can include more detail and structured formatting, but the user request here emphasizes plain paragraph format in the paper, not messaging. In practice, the generator can be trained with channel tags and can learn different styles. For multilingual SMS, abbreviations may differ by language, and excessive abbreviation can reduce comprehension. A conservative strategy is to keep messages concise but not cryptic, and to include a clear call-to-action and a staffed contact option when needed.

V. Evaluation, Safety, and Equity Across Languages

Evaluation of personalized multilingual healthcare messaging is inherently multi-dimensional. Traditional NLP metrics such as BLEU or ROUGE are weak proxies because multiple valid messages can express the same intent, and because safety and policy compliance matter more than n-gram overlap. Evaluation should instead be anchored to task outcomes, semantic fidelity to authoritative facts, and safety properties such as privacy preservation and avoidance of harmful instructions [47]. In multilingual settings, evaluation must also ensure comparable quality across languages and dialects, because uneven quality can become an equity issue.

Semantic fidelity can be evaluated by comparing the message content to the message plan derived from authoritative sources. This requires extracting entities and relations from the generated message and checking them against the plan. For example, appointment time, date, location, provider name, and preparation instructions can be extracted and validated. Numeric and temporal consistency checks are particularly important. In multilingual messages, date formats and time expressions vary, so extraction must be language-aware [48]. For some languages, numerals may be written in different scripts. Evaluation pipelines should normalize these representations before comparison.

Factuality in generative systems also includes implicit facts, such as implied certainty and implied authorization. A message that says a procedure is confirmed when it is only requested can mislead. Therefore, evaluation should include modality and hedging analysis. In multilingual contexts, modality is expressed differently, and direct translation can alter modality strength. One can evaluate modality consistency by annotating a sample of message plans with required modality and checking whether the generated message matches [49]. Automated modality classifiers can help, but they must be validated per language.

Readability and comprehensibility can be evaluated through language-specific readability measures and human comprehension tests. Many readability formulas are language-specific and may not exist for all languages. In such cases, one can use proxy measures such as sentence length, word frequency in large corpora, and the rate of technical term usage. Human evaluation remains important, especially for low-resource languages and dialects. However, human evaluation must be structured to capture misunderstandings that matter operationally, such as misinterpreting fasting instructions or arrival times.

Tone and cultural appropriateness require evaluation beyond literal correctness [50]. A message can be semantically correct but socially inappropriate. For example, certain honorifics may be required when addressing elders. Some cultures prefer indirect requests rather than imperative commands. Evaluating these properties can be done via native-speaker review panels and through complaint and satisfaction analytics after deployment. One should avoid optimizing solely for engagement metrics because they can be biased by access and trust differences. Instead, evaluation should include fairness-aware analysis that examines whether certain groups receive lower-quality messages or experience more failures [51].

Safety evaluation must focus on failure modes that are specific to healthcare messaging. A common failure mode is hallucinating additional instructions, such as suggesting a change in medication timing or interpreting symptoms. Another failure mode is disclosing sensitive information on insecure channels. A third is sending the wrong message to the wrong recipient due to identity resolution errors, which is primarily a systems issue but can be exacerbated by automation. Safety evaluation should therefore combine content-based checks with system-level checks on identity and consent enforcement.

Privacy evaluation in multilingual messaging includes testing whether the model can be induced to reveal sensitive details through prompt injection or user manipulation in inbound channels. In practice, inbound messages can include attempts to trick the system into revealing test results or appointment details [52]. The system should require authentication or use secure channels for sensitive information. Evaluation should include adversarial testing where testers attempt to elicit restricted information in multiple languages, including code-switched prompts. The system should also be tested for leakage through paraphrase, where the model might reveal sensitive content indirectly.

Bias and equity evaluation should examine disparities across languages in multiple dimensions: comprehension, error rates in entity extraction, generation fidelity, and escalation rates to humans. A system that routes low-confidence cases to humans may inadvertently create longer wait times for speakers of low-resource languages. Therefore, routing policies should be evaluated for their impact on service equity [53]. If a language consistently triggers human routing due to low confidence, the organization may need to invest in language resources or interpreter capacity. Equity evaluation should also examine whether messages differ in politeness or helpfulness across languages due to training data imbalance.

Robustness evaluation should include stress tests for noisy inputs, such as misspellings, colloquial phrases, and speech-to-text errors in transcripts. For multilingual service text, robustness must handle mixed scripts and transliteration. Evaluation datasets should include these phenomena, not only clean text. One practical method is to sample real inbound messages per language and annotate them with intents and desired actions, then measure end-to-end performance. Another is to simulate noise, but simulation must reflect real distributions, otherwise models may become robust to unrealistic noise while failing on common patterns [54].

Another key aspect is calibration and uncertainty estimation. In healthcare messaging, it is often better to abstain than to guess. Therefore, evaluation should measure not only accuracy but also the quality of confidence scores and the trade-off between automation and human routing. Calibration can be evaluated using reliability diagrams and expected calibration error per language. Because language distributions differ, one should avoid reporting a single global calibration score. Instead, per-language calibration metrics should be tracked, and thresholds should be tuned per language and per use case [55].

Human-in-the-loop evaluation is essential for safety-critical messaging. Human reviewers can evaluate whether messages are appropriate, clear, and policy-compliant. In multilingual contexts, reviewers should be native speakers familiar with healthcare communication norms. Review protocols should include structured rubrics for factual correctness, completeness, tone, and privacy. Importantly, reviewers should also record uncertainty and note when the message should have been routed to a human agent. This feedback can improve both model training and routing policy.

Offline evaluation should be complemented by online monitoring [56]. After deployment, one can track outcomes such as appointment attendance, call-back rates, and complaint rates. However, causal attribution is difficult because many factors influence these outcomes. A cautious approach uses randomized controlled experiments where feasible, but in healthcare, experiments must be ethically reviewed and must avoid differential risk. For multilingual evaluation, experiments should be stratified by language to detect disparities. Monitoring should also include automated detection of anomalous message patterns, such as sudden shifts in wording that could indicate model drift or prompt injection.

A particularly important class of evaluation is span-level provenance. For messages that include facts, the system should be able to link each fact to its source [57]. In multilingual generation, provenance can be evaluated by checking whether the source fact exists and whether the target language span corresponds semantically. This is challenging because direct string matching does not work across languages. One approach is to store a language-agnostic representation of each fact, such as a normalized slot, and then check whether the message contains a correct realization of that slot using language-specific extraction. Another approach is to generate with explicit placeholders for facts and then replace them deterministically, ensuring exactness for critical values.

Finally, evaluation must include failure response. When the system detects a potential error or a policy violation, it should fail safely [58]. Fail-safe behavior might include reverting to a minimal template, asking the user to contact a staffed line, or sending a portal notification instead of an SMS. The effectiveness of these strategies should be evaluated by measuring whether fail-safe messages still enable task completion. In multilingual settings, fail-safe messages must be available and understandable in each supported language, otherwise failure handling becomes another equity gap.

VI. Deployment, Monitoring, and Governance in Healthcare Settings

Deploying multilingual personalized messaging in healthcare requires integration with operational systems, strong governance, and continuous monitoring. The messaging component does not operate in isolation; it depends on identity resolution, consent management, scheduling systems, electronic health records, billing systems, and contact center platforms. Each integration point introduces both opportunities for richer context and risks of inconsistency. A robust deployment architecture therefore emphasizes authoritative sources, transactional validation, and clear separation between data retrieval, decision logic, and language generation [59].

A typical deployment flow begins with an event trigger, such as an upcoming appointment, a missed appointment, a completed lab test, or an inbound customer message. The system retrieves the minimal necessary context from authoritative systems, applying consent filters that determine what content is allowed for the channel. It then constructs a message plan, retrieves vetted phrasing, and invokes controlled generation if needed. Before dispatch, it runs verification checks, logs the decision and provenance, and then sends via the approved channel. If checks fail, it routes to human review or sends a fallback message. Each step should be instrumented so that errors can be traced [60].

Identity and consent management are foundational. Messaging should be tied to verified contact points, and consent should be recorded with timestamp, channel, and language. In multilingual settings, consent capture should be available in the patient's language, and the system should handle situations where a caregiver is the contact person. Caregiver messaging introduces additional privacy considerations because the caregiver may not be authorized to receive all details. Therefore, deployment should support role-based consent and content filtering that can restrict messages to administrative content when sent to caregivers.

Audit logging is essential for accountability. Logs should record what facts were used, what message was generated or selected, what verifications were applied, and what routing decisions were made [61]. In multilingual generation, logs should include the language tag, model version, template version, and any adaptation parameters such as tone and reading level. However, logs themselves can contain sensitive content. Therefore, logging should be designed with data minimization, potentially storing message hashes, structured slots, and redacted content rather than full raw text, while still enabling investigation. Secure access controls and retention policies are also necessary.

Model updates and versioning require careful governance. Even small changes in a multilingual model can have uneven effects across languages [62]. A model update that improves English performance may degrade another language due to interference. Therefore, deployment should include per-language regression testing and canary releases. A canary can route a small fraction of messages in a given language to the new model while monitoring key metrics and complaint signals. Rollback mechanisms must be immediate. In healthcare, it is safer to adopt conservative release cycles with thorough review rather than frequent updates.

Monitoring should include both technical and operational indicators. Technical indicators include verification failure rates, language identification errors, entity mismatch rates, and distributional shifts in input text embeddings [63]. Operational indicators include call-back rates, appointment attendance, message delivery failures, opt-out rates, and complaint categories. In multilingual settings, monitoring dashboards should segment by language and site, because shifts may occur locally. For example, a new clinic may serve a different language mix, and the model may encounter unfamiliar dialects. Monitoring should also detect sudden spikes in specific phrase usage, which can indicate a prompt injection attack if a generation system is driven by inbound text.

Governance workflows should define which message types are eligible for automation. Some message categories, such as appointment reminders and administrative updates, are often suitable for high automation with deterministic content. Other categories, such as discussing test results, may require clinician involvement [64]. Governance should specify allowed content types per channel and per language, because some channels are more secure than others. The system should encode these rules so that the generator cannot override them. When governance policies change, the system should be able to update rules without retraining models, which argues for a policy layer external to the model.

Human oversight is not only a safety mechanism but also a resource planning issue. If multilingual automation routes many cases to humans due to low confidence, the organization must staff appropriately, including bilingual agents or interpreter services. Governance should therefore include capacity-aware routing, where the system chooses between different safe options depending on staffing availability [65]. For example, if interpreter capacity is limited, the system might send a minimal bilingual message directing the patient to a portal rather than initiating a complex conversation. Such decisions should be documented and reviewed because they affect access.

Security considerations include preventing unauthorized access and preventing manipulation through inbound channels. If the system uses large language models with prompts that include retrieved context, prompt injection can cause the model to disregard policy constraints. Therefore, the generation component should be isolated from raw inbound text, and inbound text should be parsed and normalized before being used as conditioning. Additionally, sensitive facts should be inserted deterministically rather than being left for the model to reproduce. Where models must handle user-provided text, one can apply sanitization and intent extraction to convert it into structured variables, reducing the risk that user text directly shapes generation [66].

Localization and language operations are ongoing tasks. Languages evolve, and organizations may expand into new regions. Deployment should therefore include a language onboarding process that defines quality thresholds, human review protocols, and fallback behavior for each language. It should also include processes for updating lexicons, such as clinic names, provider titles, and local transportation references. In multilingual environments, even the spelling of clinic names may vary by language, and inconsistencies can confuse patients. Maintaining a multilingual terminology database that is used by both retrieval and generation can improve consistency [67].

Another practical challenge is handling outbound deliverability and channel constraints. SMS delivery can fail due to carrier issues, and certain characters may be unsupported in some networks. For languages using non-Latin scripts, SMS may be encoded differently, reducing character limits. The generation system should be aware of these constraints and should adapt message length accordingly. It should also avoid characters that can be misrendered. For languages that require more characters to express the same content, the system may need to prioritize essential facts and provide links or call options for additional details.

Measuring real-world impact requires careful interpretation [68]. Changes in no-show rates or call volume can reflect external factors. Moreover, improvements for one language group can coincide with declines for another if resources are reallocated. Governance should therefore include periodic equity reviews that examine outcomes by language, age group, and other relevant categories, while respecting privacy and avoiding inappropriate profiling. These reviews can guide investments in language resources and improvements in model performance where disparities are detected.

Finally, documentation and training are part of deployment. Staff need to understand how the system behaves, what it can and cannot do, and how to intervene when it fails [69]. Patients also benefit from transparency, such as knowing that messages are automated and how to request a human. In multilingual contexts, transparency statements should be available in the patient's language. Clear communication about opt-out and preference update mechanisms can reduce frustration and improve trust.

VII. Conclusion

Personalized healthcare customer messaging using multilingual clinical and service text is best approached as a constrained, multi-component NLP system rather than a single generative model. The core technical requirement is to transform heterogeneous sources into communication-appropriate facts and strategies while enforcing privacy, safety, and policy constraints. Multilinguality adds complexity because correctness includes pragmatic and cultural adequacy in addition to semantic fidelity, and because language coverage and quality can become equity concerns when uneven. Effective systems therefore combine deterministic planning from authoritative sources, multilingual retrieval of vetted phrasing, controlled generation for adaptation, and rigorous verification that checks entity consistency, modality, and channel-appropriate disclosure [70].

Personalization is valuable when it improves clarity, reduces friction, and respects consent, but it introduces risk when it surfaces sensitive attributes or relies on biased historical signals. A conservative personalization approach emphasizes explicit preferences, uncertainty-aware updates, and routing to humans when confidence is low. Evaluation must match this reality by measuring semantic fidelity to message plans, privacy and safety violations, readability, tone appropriateness, and disparities across languages and dialects. Offline testing should be complemented by online monitoring with per-language segmentation, canary releases, and rapid rollback capabilities.

Deployment in healthcare requires governance that defines which message categories can be automated, how consent is enforced, how audits are conducted, and how human oversight is integrated. Practical success depends on robust integration with operational systems, careful handling of identity and channel constraints, and sustained language operations for terminology and localization. Within these constraints, multilingual NLP can support clearer and more accessible communication, provided that system design prioritizes provenance, verification, and equitable service quality across languages rather than maximizing generative flexibility [71].

REFERENCES

[1] B. Büyükkaya, S. Altındiş, and G. Ünlü, "Attitudes and familiarity of hospital administrative and medical staff towards artificial intelligence: A cross-sectional study in turkey," *Asia Pacific Journal of Health Management*, 5 2025.

[2] "Thomson corp.," *Mergent's Dividend Achievers*, vol. 1, pp. 332–332, 10 2004.

[3] I. B. Majele and N. W. Mwaikambo, "Evaluating customers' perceptions on the reliability of healthcare services provided by government hospitals: A case study of ilala district council," *Asian Journal of Advanced Research and Reports*, vol. 18, pp. 182–191, 10 2024.

[4] A. Mittal and A. Mantri, "A literature survey on healthcare supply chain management," *F1000Research*, vol. 12, pp. 759–759, 6 2023.

[5] S. H. Kukkuhalli, "Enabling customer 360 view and customer touchpoint tracking across digital and non-digital channels," *Journal of Marketing & Supply Chain Management*, vol. 1, no. 3, 2022.

[6] M. Anshari, "Redefining electronic health records (ehr) and electronic medical records (emr) to promote patient empowerment," *IJID (International Journal on Informatics for Development)*, vol. 8, pp. 35–39, 6 2019.

[7] I. M. Shor, “Healthcare development based on public-private partnership: Opportunities and limitations,” *Vestnik of North Ossetian State University*, pp. 182–189, 3 2022.

[8] N. Trakulmaykee, C. Choksuchat, K. W. nguan Jetwanna, and K. S. Inthanuchit, “Moderating roles of user’s intention to use line official account in healthcare context: body mass index,” *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 14, pp. 6807–6807, 12 2024.

[9] F. Schiavone, D. Leone, A. Sorrentino, and A. Scaletti, “Re-designing the service experience in the value co-creation process: an exploratory study of a health-care network,” *Business Process Management Journal*, vol. 26, pp. 889–908, 6 2020.

[10] P. Mechinda and P. G. Patterson, “The impact of service climate and service provider personality on employees’ customer-oriented behavior in a high-contact setting,” *Journal of Services Marketing*, vol. 25, pp. 101–113, 4 2011.

[11] L. L. Dijkstra and J. H. van der Bij, “Quality function deployment in healthcare. methods for meeting customer requirements in redesign and renewal,” *International Journal of Quality & Reliability Management*, vol. 19, pp. 67–89, 2 2002.

[12] T. T. Borishade, R. E. Worlu, O. Kehinde, A. S. Ibidunni, O. O. Ogunnaike, J. Dirisu, O. Ogueyungbo, and F. Peter, “A dataset of customer loyalty and variation in perception of customers across demographic characteristics in healthcare sector of nigeria.,” *Data in brief*, vol. 20, pp. 353–357, 8 2018.

[13] T. E. Grissom, A. DuKatz, H. Kordylewski, and R. P. Dutton, “Bring out your data: The evolution of the national anesthesia clinical outcomes registry (nacor),” *International Journal of Computational Models and Algorithms in Medicine*, vol. 2, pp. 51–69, 4 2011.

[14] R. Prasetya and D. Santoso, “Instrumental variables in high-dimensional settings: Estimation and inference under sparsity and endogeneity constraints,” *International Journal of Theoretical, Computational, and Applied Multidisciplinary Sciences*, vol. 8, no. 9, pp. 1–22, 2024.

[15] R. A. da Silva, M. C. da Silva, and M. M. C. dos Santos, “Empreendedorismo na enfermagem: A atuaÇÃo do enfermeiro em consultorias,” *Revista ft*, vol. 29, pp. 47–48, 10 2025.

[16] O. Amankwah, W. W. Choong, and A. H. Mohammed, “Modelling the influence of healthcare facilities management service quality on patients satisfaction,” *Journal of Facilities Management*, vol. 17, pp. 267–283, 7 2019.

[17] C. P. Price and A. S. John, “The real value of laboratory medicine,” *The journal of applied laboratory medicine*, vol. 1, pp. 101–103, 7 2016.

[18] C. Huichao, X. Luo, Z.-L. Zhang, and Q. Zhou, “Stochastic bi-level programming model for home healthcare scheduling problems considering the degree of satisfaction with visit time,” *Journal of Systems Science and Systems Engineering*, vol. 30, pp. 572–599, 9 2021.

[19] T. Hilken, J. Heller, and D. Mahr, “Closing the customer imagination gap with augmented and virtual reality,” *NIM Marketing Intelligence Review*, vol. 15, pp. 30–35, 10 2023.

[20] J. R. Machireddy, “Data quality management and performance optimization for enterprise-scale etl pipelines in modern analytical ecosystems,” *Journal of Data Science, Predictive Analytics, and Big Data Applications*, vol. 8, no. 7, pp. 1–26, 2023.

[21] D. J. Scotti, J. Harmon, and S. J. Behson, “Structural relationships between work environment and service quality perceptions as a function of customer contact intensity: implications for human service strategy.,” *Journal of health and human services administration*, vol. 32, pp. 195–234, 6 2009.

[22] A. Helkkula, E. Arnould, and A. Chen, “Glocalization in service cultures: Tensions in customers’ service expectations and experiences,” *Journal of Service Research*, vol. 26, pp. 233–250, 5 2022.

[23] “Sysco corp.,” *Mergent’s Dividend Achievers*, vol. 5, pp. 271–271, 1 2008.

[24] X. Rao, L. Luo, J. Xiang, and X. Wang, “The impact of perceived value, customer expectations, and patient experience on the satisfaction of contracted patients in hospitals.,” *BMC health services research*, vol. 25, pp. 7–, 1 2025.

[25] O. A. Olaitan, “Evaluating strategic planning as intervention tool for improving healthcare delivery in private hospitals in ojodu, lagos,” *Texila International Journal of Management*, vol. 8, pp. 137–143, 8 2022.

[26] D. Das and A. Sengar, “A fuzzy analytic hierarchy process-based analysis for prioritization of barriers to the adoption of ehealth in india.,” *International journal of medical informatics*, vol. 165, pp. 104830–104830, 7 2022.

[27] P. A. Olujimi and A. Ade-Ibijola, “Nlp techniques for automating responses to customer queries: a systematic review,” *Discover Artificial Intelligence*, vol. 3, 5 2023.

[28] J. R. Machireddy, “Integrating predictive modeling with policy interventions to address fraud, waste, and abuse (fwa) in us healthcare systems,” *Advances in Computational Systems, Algorithms, and Emerging Technologies*, vol. 7, no. 1, pp. 35–65, 2022.

[29] M. Vucha, “High speed cryptography architecture for health information exchange,” *International Journal of Advanced Trends in Computer Science and Engineering*, pp. 1443–1448, 8 2019.

[30] L. C. Burr, R. A. Patterson, E. Rolland, and K. Ward, “Integration of e-crm in healthcare services: A frame-

work for analysis," *International Journal of E-Business Research*, vol. 3, pp. 1–12, 4 2007.

[31] C. Yildirim, M. Demir, and A. Kaya, "A comparative analysis of fiscal investments in preventive healthcare measures and their effects on national economic productivity," *Advanced Studies in Human Knowledge, Innovation, and Societal Development*, vol. 15, no. 8, 2025.

[32] S. R. Shetty, D. Chatterjee, and Y. P. Pai, "Performance of third party administrators in indian healthcare," *Journal of Health Management*, vol. 24, pp. 652–661, 10 2022.

[33] S. Agarwal and R. Singh, "Customers' perception towards accountability of diagnostic centres: Evidence from india.," *Journal of multidisciplinary healthcare*, vol. 16, pp. 2947–2961, 10 2023.

[34] D. V. Dimitrov, "Current healthcare in bulgaria: time for predictive diagnostics and preventive medicine.," *The EPMA journal*, vol. 1, pp. 607–610, 10 2010.

[35] M. A. Hadiyat, R. D. Wahyudi, and Y. Sari, "Survival analysis for customer satisfaction: A case study," *IOP Conference Series: Materials Science and Engineering*, vol. 273, pp. 012029–, 11 2017.

[36] S. H. Kukkuhalli, "Improving digital sales through reducing friction points in the customer digital journey using data engineering and machine learning," *International Journal of Innovative Research in Engineering & Multidisciplinary Physical Sciences*, vol. 10, no. 3, 2022.

[37] P. Sanan, "Crm capability and service innovation in healthcare: Evidence from india," *Asia Pacific Journal of Health Management*, 4 2023.

[38] G. SARMA and B. C. BARUA, "Direct-to-consumer pharmaceuticals as a catalyst for employment growth in kamrup rural : Analysing infrastructure and consumer behaviour," *Indian Commerce Association*, vol. 78, pp. 176–190, 1 2025.

[39] L. Golemanov, K. Zgodavova, and A. Bourek, "Multilingual virtual quality center providing e-learning facilities for e-health and modern healthcare management education and training," *Acta Mechanica Slovaca*, vol. 14, pp. 6–13, 10 2010.

[40] C. Avadhani, D. G. Udayasuriyan, and P. K. V. Rao, "Common health hazards in healthcare facilities & impact on healthcare workers/professionals/patients – remedial measures – an analysis," *International Journal For Multidisciplinary Research*, vol. 5, 1 2023.

[41] J. W. Gardner, K. Linderman, and K. L. McFadden, "Managing quality crossroads in healthcare: An integrative supply chain perspective," *Quality Management Journal*, vol. 25, pp. 2–17, 1 2018.

[42] E. Mooi, V. Kashyap, and M. van Aken, "Governance and customer value creation in business solutions," *Journal of Business & Industrial Marketing*, vol. 35, pp. 1089–1098, 3 2020.

[43] "Air products & chemicals, inc.," *Mergent's Dividend Achievers*, vol. 5, pp. 4–4, 1 2008.

[44] Y. Yang, "Does economic growth induce smoking?—evidence from china," *Empirical Economics*, vol. 63, no. 2, pp. 821–845, 2022.

[45] S. Alshawi, F. Missi, and T. Eldabi, "Healthcare information management: the integration of patients' data," *Logistics Information Management*, vol. 16, pp. 286–295, 6 2003.

[46] W. Aloulou, "Untangling the effects of supply chain integration and digital transformation on healthcare service providers' performance: resilience to disruptions as mediator," *Journal of Asia Business Studies*, vol. 19, pp. 1040–1061, 6 2025.

[47] "Owens & minor, inc.," *Mergent's Dividend Achievers*, vol. 5, pp. 206–206, 10 2008.

[48] T. Mettler, "Transformation of the hospital supply chain: How to measure the maturity of supplier relationship management systems in hospitals?," *International Journal of Healthcare Information Systems and Informatics*, vol. 6, pp. 1–13, 4 2011.

[49] C. Coenen, D. Waldburger, and D. von Felten, "Fm servicebarometer: monitoring customer perception of service performance," *Journal of Facilities Management*, vol. 11, pp. 266–278, 7 2013.

[50] M. A. Aziz and D. Nawaz, "Total quality m anagement a pproach in solid waste of h ealthcare system in pakistan," *Journal of Social Research Development*, vol. 2, pp. 201–213, 12 2021.

[51] P. I. Oja, T. Kouri, and A. Pakarinen, "Health centres' view of the services provided by a university hospital laboratory: Use of satisfaction surveys," *Scandinavian journal of primary health care*, vol. 28, pp. 24–28, 3 2010.

[52] J. R. Machireddy, "Automation in healthcare claims processing: Enhancing efficiency and accuracy," *International Journal of Science and Research Archive*, vol. 09, no. 01, pp. 825–834, 2023.

[53] Y.-F. Chuang, S.-H. Chia, and J. Y. Wong, "Customer value assessment of pharmaceutical marketing in taiwan," *Industrial Management & Data Systems*, vol. 113, pp. 1315–1333, 9 2013.

[54] P. A. Danarahmanto, S. Supriyadi, and J. F. Pratami, "Kepuasan pasien berbasis inovasi pengalaman pada fasilitas pelayanan kesehatan gigi dan mulut di jawa barat," *Journal of Management and Business Review*, vol. 19, pp. 108–122, 7 2022.

[55] G. Kurniawan and B. Rahmanto, "A computational analysis of socioeconomic factors driving health disparities between urban and rural populations in industrialized economies," *International Journal of Theoretical, Computational, and Applied Multidisciplinary Sciences*, vol. 9, no. 4, pp. 1–22, 2025.

[56] C. Selvamuthu, H. H. Murugadoss, H. Murugadoss, S. Padmaja, V. Sri, E. Jayasri, and M. A. Varshni,

"Strengthening india's healthcare system: Combating fraud with technology and policy reforms," *Journal of Pioneering Medical Sciences*, vol. 14, pp. 92–104, 8 2025.

[57] L. HAMADE, "Customer relationship management and its impact on patient satisfaction: A quantitative study in lebanon," *BUSINESS EXCELLENCE AND MANAGEMENT*, vol. 12, pp. 35–46, 6 2022.

[58] L. Setiawan and K. Keni, "The influence of price perception, customer trust, and customer experience on purchasing decisions at mtm pharmacy in bekasi," *International Journal of Application on Economics and Business*, vol. 1, pp. 2727–2735, 11 2023.

[59] A. L. Popa, N. N. Ţarcă, D. V. Sasu, S. A. Bodog, R. D. Roşca, and T. M. Tarcza, "Exploring marketing insights for healthcare: Trends and perspectives based on literature investigation," *Sustainability*, vol. 14, pp. 10499–10499, 8 2022.

[60] I. bin Zakaria, M. Z. bin Mohd Salleh, and I. N. A. M. F. Kozako, "The impact of social media marketing on brand loyalty among generation y towards luxury fashion brands," *Malaysian Journal of Business and Economics (MJBE)*, vol. 6, pp. 25–25, 12 2019.

[61] S. Yeshaw, G. Asrade, A. Hagos, M. G. Chanie, and N. Worku, "Practice of balanced scorecard implementation and its contributing factors among public primary hospital professionals in central gondar zone, northwest ethiopia.," *Frontiers in public health*, vol. 12, pp. 1424133–, 1 2025.

[62] M. Paredes and A. Villalobos, "Economic recessions, technology-enabled mental health services, and healthcare expenditures in developed nations," *Annual Compendium of Multidisciplinary Scientific Insights and Technological Frontiers*, vol. 15, no. 7, pp. 1–10, 2025.

[63] A. Widtfeldt and J. R. Widtfeldt, "Total quality management in american industry.," *AAOHN journal : official journal of the American Association of Occupational Health Nurses*, vol. 40, pp. 311–318, 6 1992.

[64] P. Lei and A. Jolibert, "A three-model comparison of the relationship between quality, satisfaction and loyalty: an empirical study of the chinese healthcare system," *BMC health services research*, vol. 12, pp. 436–436, 11 2012.

[65] K. Jooste, "Current nursing practice: challenges and successes," *Journal of nursing management*, vol. 19, pp. 833–836, 9 2011.

[66] N. Salsabila and S. ., "The impact of perceived ease of use and customer trust on service attractiveness: Investigating the moderating effect of age in digital healthcare service," *Journal of Economics, Finance And Management Studies*, vol. 7, 12 2024.

[67] C. Challinor, "Training and facilitating the use of reminiscence resources in a health promotion library context.," *Health information and libraries journal*, vol. 42, pp. 578–582, 9 2025.

[68] S. H. Kukkuhalli, "Increasing digital sales revenue through 1:1 hyper-personalization with the use of machine learning for b2c enterprises," *Journal of Artificial Intelligence, Machine Learning and Data Science*, vol. 1, no. 1, 2023.

[69] K. Yeganegi, "The integration of qfd technique and value engineering and its applying in a healthcare center," *SSRN Electronic Journal*, 1 2011.

[70] K. Olesen and H. Bathula, "A meta-analysis of the determinants of patient satisfaction and loyalty.," *Health marketing quarterly*, vol. 39, pp. 191–210, 3 2022.

[71] S. N. Faijah and N. Nurhayati, "Exploring customer satisfaction in addition apotek services in banjarbaru: The role of product involvement, brand familiarity, and consumer attitudes," *Eduvest - Journal of Universal Studies*, vol. 5, pp. 11684–11698, 9 2025.