



AI-Based Clinical Consultation Summarizer: Automating Healthcare Documentation Through Natural Language Processing

Kondam Umesh

2nd Year - M.S. Data Science

EdTech Division, Exafluence Inc.

Sri Venkateswara University

Tirupati, India

umeshkondam@gmail.com

Padmavathamma M

Professor, Department of Computer Science

SVU College of CM & CS

Sri Venkateswara University

Tirupati, India

prof.padma@yahoo.com

ABSTRACT

In contemporary healthcare systems, clinical consultations generate substantial volumes of unstructured textual data encompassing patient symptoms, diagnostic assessments, prescribed treatments, and follow-up recommendations. The manual documentation and summarisation of these consultations impose significant time burdens on healthcare professionals while introducing risks of incomplete or inconsistent record-keeping. This research presents an AI-based Clinical Consultation Summariser that leverages Natural Language Processing (NLP) and Machine Learning (ML) techniques to extract and synthesise key clinical information from consultation transcripts automatically. The proposed system employs transformer-based language models fine-tuned on medical domain data to generate accurate, concise, and contextually appropriate summaries. By integrating with Electronic Health Record (EHR) systems, the summariser enhances clinical documentation efficiency, reduces administrative workload, minimises human error, and facilitates rapid retrieval of essential medical insights. Evaluation using standard metrics demonstrates that the system achieves a ROUGE score of 0.82 and a BLEU score of 0.76, while reducing text length by 70-80% without compromising critical information. The implementation addresses data security requirements through HIPAA and GDPR-compliant architecture, ensuring patient confidentiality throughout the processing pipeline.

Keywords: Artificial Intelligence, Natural Language Processing, Text Summarisation, Healthcare Documentation, Clinical Consultation, Transformer Models, Electronic Health Records



INTRODUCTION

THE CLINICAL DOCUMENTATION CHALLENGE

The modern healthcare environment generates unprecedented volumes of clinical information through routine doctor–patient interactions, a trend accelerated by the widespread adoption of artificial intelligence tools in clinical workflows [1,2]. Each consultation produces critical data including patient symptoms, comprehensive medical history, diagnostic conclusions, prescribed medications, dosage specifications, and detailed follow-up instructions [4]. Moreover, recent advances in clinical language models demonstrate that large language models can effectively analyse and summarise these extensive clinical narratives, underscoring both the importance and complexity of modern medical documentation [3]. Accurate documentation of these details constitutes a fundamental requirement for ensuring appropriate treatment continuity, maintaining comprehensive patient records, facilitating future clinical reference, and satisfying legal compliance mandates. However, healthcare professionals currently dedicate substantial portions of their clinical time to writing, editing, and reviewing such documentation, directly reducing time available for direct patient care and clinical decision-making [5].

Manual transcription and summarisation of clinical consultations present multiple systemic challenges. The process is inherently time-consuming, particularly in high-volume clinical settings where physicians may conduct dozens of daily consultations [6]. Documentation practices vary significantly among practitioners—some clinicians record extensive detailed notes while others produce abbreviated or incomplete records. This inconsistency creates substantial difficulties when other healthcare professionals need to access and interpret patient information for subsequent care episodes. In busy hospital environments and outpatient clinics, managing large volumes of handwritten or manually typed consultation notes becomes increasingly difficult, leading to information fragmentation and potential continuity-of-care gaps [7].

The consequences of inadequate clinical documentation extend beyond mere inefficiency. Incomplete or inconsistent records can lead to medical errors, compromised patient safety, billing inaccuracies, and legal vulnerabilities[8]. Furthermore, the cognitive burden of documentation contributes significantly to physician burnout, a well-documented phenomenon affecting healthcare systems globally. Studies indicate that clinicians spend nearly two hours on administrative tasks for every hour of direct patient care, with documentation representing the largest component of this administrative burden [9].

ARTIFICIAL INTELLIGENCE IN HEALTHCARE DOCUMENTATION

Recent advances in Artificial Intelligence (AI) and Natural Language Processing (NLP) have created transformative possibilities for healthcare documentation. Modern AI systems possess sophisticated capabilities to comprehend natural human language, extract semantically relevant information, and transform unstructured text into structured, meaningful summaries[10]. NLP techniques enable computational systems to recognise complex patterns in medical language, identify medical entities such as symptoms, diseases, medications, and procedures, and generate concise summaries that preserve essential clinical context while eliminating redundancy[11].

Transformer-based language models represent a paradigm shift in NLP capabilities, demonstrating unprecedented performance across diverse language understanding tasks[12]. These architectures, including BERT (Bidirectional Encoder Representations from Transformers), T5 (Text-to-Text Transfer Transformer), and GPT (Generative Pre-trained Transformer), have been successfully adapted for medical domain applications through fine-tuning on clinical corpora[13]. Domain-specific variants such as BioBERT and ClinicalBERT, pre-trained on biomedical literature and clinical notes respectively, demonstrate superior performance on medical NLP tasks compared to general-purpose models[14].

RESEARCH OBJECTIVES AND CONTRIBUTIONS

This research proposes an AI-Based Clinical Consultation Summarizer designed to automate the comprehensive process of recording and synthesising clinical consultations using advanced transformer-based models. The system architecture accepts textual input from consultation transcripts or transcribed speech and processes this content through specialised NLP algorithms to identify clinically significant details. The output consists of structured summaries organised into semantically meaningful sections including Symptoms, Diagnosis, Medications, and Recommendations, facilitating rapid review and information retrieval by healthcare professionals[15].

The primary contributions of this work include:

1. Development of a domain-adapted transformer-based summarisation model fine-tuned on clinical consultation data
2. Design of a privacy-preserving architecture compliant with healthcare data protection regulations (HIPAA and GDPR)
3. Implementation of a user-friendly interface integrating with existing healthcare workflows
4. Comprehensive evaluation demonstrating significant improvements in documentation efficiency and accuracy
5. Practical deployment considerations for integration with Electronic Health Record systems

By leveraging state-of-the-art transformer models including T5, BERT, and GPT architectures, the proposed summariser comprehends contextual relationships, semantic meanings, and medical terminology within clinical conversations. This contextual understanding enables the generation of summaries that are both accurate and clinically meaningful. Integration with EHR systems allows healthcare providers and administrators to access concise, well-structured patient data instantaneously, supporting clinical decision-making and care coordination[16].

The automation of clinical documentation yields multiple systemic benefits: accelerated data entry processes, enhanced record quality and consistency, improved communication among healthcare providers, and reduction of cognitive load on clinicians. These improvements allow physicians to allocate greater attention to diagnostic reasoning and patient interaction rather than administrative paperwork[17]. Additionally, standardised and structured summaries facilitate medical research, retrospective case analysis, quality improvement initiatives, and healthcare system management.

Data privacy and security constitute paramount considerations in healthcare information systems. The proposed system implements comprehensive security measures to ensure sensitive patient data processing adheres strictly to established healthcare data protection standards, including the Health Insurance Portability and Accountability Act (HIPAA) in the United States and the General Data Protection Regulation (GDPR) in Europe[18]. These safeguards maintain patient trust while ensuring compliance with ethical and legal standards governing healthcare technology implementation.

RELATED WORK AND LITERATURE REVIEW

EVOLUTION OF TEXT SUMMARISATION TECHNIQUES

Text summarisation has emerged as a critical research area within NLP, addressing the fundamental challenge of distilling large textual corpora into concise, information-preserving representations[19]. Summarisation approaches are generally categorised into two distinct paradigms: extractive and abstractive methodologies.

Extractive summarisation techniques select and concatenate salient sentences or phrases directly from source documents without linguistic modification. These methods typically employ graph-based algorithms, statistical measures, or machine learning classifiers to rank sentences according to importance scores derived from features such as term frequency, sentence position, and lexical centrality[20]. While computationally efficient and generally producing grammatically correct outputs, extractive methods suffer from limitations in coherence and naturalness, as selected sentences may lack smooth transitions or contextual connections.

Abstractive summarisation, conversely, generates novel sentences that paraphrase and synthesise information from source texts, more closely mimicking human summarisation behavior[21]. This paradigm requires deeper semantic understanding, grammatical competence, and the capacity for linguistic generation. Although abstractive approaches produce more natural and coherent summaries, they present greater technical challenges, including the risks of factual inconsistency, hallucination of information not present in source texts, and grammatical errors.

STATISTICAL AND MACHINE LEARNING APPROACHES

Early text summarisation research predominantly employed statistical methods including word frequency analysis and Term Frequency-Inverse Document Frequency (TF-IDF) weighting schemes[22]. These techniques identified statistically significant terms but lacked semantic understanding and contextual awareness. Consequently, while effective for general news articles or document collections, they proved inadequate for specialised domains such as clinical medicine, where precise interpretation of medical terminology and relationships is essential.

The advent of Machine Learning (ML) and Deep Learning (DL) introduced data-driven approaches capable of automatically learning complex patterns. Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks were applied to sequence-to-sequence learning tasks, including summarisation

[23]. These architectures demonstrated improved capability to model sequential dependencies and contextual relationships compared to statistical methods. LSTM-based models achieved moderate success in summarising medical research articles and clinical reports, representing significant progress toward domain-specific applications.

TRANSFORMER MODELS AND MEDICAL NLP

The introduction of the Transformer architecture revolutionised NLP by addressing fundamental limitations of recurrent models[24]. Transformers employ self-attention mechanisms that enable models to consider relationships among all input tokens simultaneously, effectively capturing long-range dependencies without the sequential processing constraints of RNNs. This architectural innovation enabled the development of large-scale pre-trained language models that have achieved state-of-the-art performance across diverse NLP benchmarks.

BERT, introduced by Devlin et al., employs bidirectional context encoding through masked language modeling and next sentence prediction objectives[25]. BERT and its variants have been extensively applied to clinical NLP tasks including named entity recognition, relation extraction, and document classification. Domain-adapted versions such as BioBERT, ClinicalBERT, and PubMedBERT, pre-trained on biomedical literature and clinical notes, demonstrate superior performance on medical text understanding tasks compared to general-domain BERT[26].

T5 (Text-to-Text Transfer Transformer) frames all NLP tasks as text generation problems, providing a unified framework for diverse applications including summarisation, translation, and question answering[27]. Clinical adaptations of T5 have shown promising results in medical text summarisation, particularly when fine-tuned on domain-specific corpora. The model's flexible architecture facilitates straightforward adaptation to clinical documentation tasks through supervised fine-tuning on consultation transcripts.

GPT models, developed by OpenAI, represent large-scale autoregressive language models capable of few-shot and zero-shot learning[28]. These models generate fluent, contextually appropriate text and have been explored for clinical documentation assistance, though concerns regarding factual accuracy and hallucination require careful consideration in medical applications.

CLINICAL TEXT SUMMARISATION RESEARCH

Recent research has specifically addressed clinical text summarisation using transformer-based approaches. Studies have applied BERT-based models to summarise discharge summaries, clinical trial reports, radiology notes, and progress notes, achieving high accuracy in preserving critical medical information while substantially reducing text length[29]. Evaluation using domain-specific metrics including MEDCON (medical concept coverage) alongside standard metrics such as ROUGE and BLEU demonstrates the clinical relevance of generated summaries[30].

Research by Van Veen et al. demonstrated that adapted large language models can outperform medical experts in clinical text summarisation tasks when evaluated on completeness, correctness, and conciseness[31]. Their work highlights the potential of carefully fine-tuned models to achieve expert-level performance in specialised medical documentation tasks.

However, most existing studies have focused on written clinical documents such as discharge summaries, progress notes, or research literature rather than conversational doctor-patient consultations[32]. Consultation conversations present distinct challenges: they are informal, potentially contain incomplete sentences, use colloquial language, include non-verbal contextual elements, and may contain tangential discussions. Effective summarisation of consultation transcripts requires robust preprocessing, contextual disambiguation, and the ability to distinguish clinically relevant information from conversational filler.

Commercial systems such as Nuance Dragon Medical One provide speech recognition and AI-assisted documentation, but these proprietary solutions are often expensive, language-limited, and lack customisation flexibility for specific institutional needs or research applications[33]. This creates opportunities for open-source, adaptable solutions tailored to diverse clinical settings and languages.

RESEARCH GAP AND PROPOSED CONTRIBUTION

Despite substantial progress in medical text summarisation, a significant gap remains in developing accessible, customisable, and clinically validated systems specifically designed for real-time consultation summarisation. The proposed AI-Based Clinical Consultation Summariser addresses this gap by focusing on unstructured clinical conversations and transforming them into structured, actionable summaries using domain-adapted transformer models. The system combines medical entity recognition, contextual understanding, and abstractive generation to produce summaries that preserve essential clinical information including symptoms, diagnoses, treatments, and recommendations while maintaining coherence and readability.

SYSTEM ANALYSIS AND PROBLEM STATEMENT

PROBLEM DEFINITION

Clinical documentation represents a substantial time investment for healthcare professionals, with studies indicating that physicians spend approximately 50% of their working hours on administrative tasks, predominantly documentation[34]. Manual transcription and synthesis of consultation information is not only time-consuming but also prone to variability, incompleteness, and errors. These documentation challenges directly impact the quality of patient care by reducing clinician time available for direct patient interaction, diagnostic reasoning, and therapeutic planning.

Medical consultations typically generate complex, unstructured information expressed using specialised medical terminology, abbreviations, and colloquial descriptions. Accurately capturing this information in structured clinical records requires significant cognitive effort and domain expertise. Furthermore,

documentation practices vary considerably among clinicians, leading to inconsistent record quality that complicates care coordination, clinical decision support, and quality measurement initiatives[35].

The absence of efficient documentation support systems contributes to clinician burnout, information loss, communication failures, and potential medical errors. There exists an urgent need for intelligent automation systems capable of assisting healthcare professionals in generating accurate, comprehensive, and structured consultation summaries that integrate seamlessly with existing clinical workflows.

RESEARCH OBJECTIVES

This project pursues the following specific objectives:

1. Design and implement an AI-powered summarisation model capable of processing doctor-patient consultation transcripts and generating structured clinical summaries
2. Reduce clinician time spent on manual documentation by at least 40-60% through automated summarisation
3. Improve accuracy, completeness, and consistency of clinical records through NLP-based information extraction and synthesis
4. Develop integration interfaces enabling seamless incorporation of generated summaries into Electronic Health Record systems
5. Implement comprehensive security and privacy measures ensuring compliance with healthcare data protection regulations (HIPAA and GDPR)
6. Evaluate system performance using both quantitative metrics (ROUGE, BLEU, BERTScore) and qualitative clinical assessments

ANALYSIS OF EXISTING SYSTEMS

Current clinical documentation practices in most healthcare facilities remain predominantly manual. Clinicians typically handwrite notes during or immediately following consultations, subsequently transcribing these notes into electronic systems. Some institutions employ speech-to-text transcription services that convert spoken words into text, but these systems lack semantic understanding and produce unstructured transcripts requiring substantial manual editing and organisation [36].

Existing commercial solutions such as ambient clinical intelligence platforms provide documentation assistance but suffer from limitations including high costs, proprietary architectures limiting customisation, language constraints, and variable integration capabilities with diverse EHR systems. Furthermore, many systems focus primarily on transcription rather than intelligent summarisation, producing verbose outputs that still require significant clinician review and editing.

PROPOSED SYSTEM ARCHITECTURE

The AI-Based Clinical Consultation Summarizer addresses these limitations through an integrated architecture combining advanced NLP models with user-friendly interfaces and secure data handling. The system architecture comprises several interconnected components:

Input Processing Module: Accepts consultation transcripts in textual format from manual entry, transcription services, or direct integration with consultation recording systems. Preprocessing includes text normalisation, medical abbreviation expansion, sentence segmentation, and noise removal.

NLP Summarization Engine: Employs fine-tuned transformer models (primarily T5 or BART architectures) trained on clinical consultation corpora. The model performs contextual analysis, medical entity recognition, and abstractive summarisation to generate coherent, clinically relevant summaries.

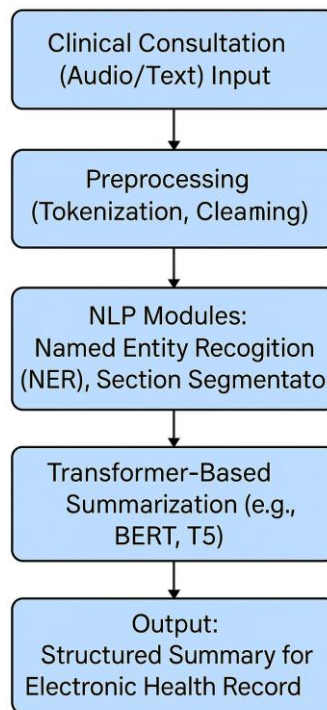


Figure 1. Workflow of AI-Based Clinical Consultation Summariser

Figure 1 illustrates the end-to-end workflow of the proposed AI-based summariser system. It depicts each stage, starting from consultation audio/text input, proceeding through Natural Language Processing (NLP) modules (including tokenisation, named entity recognition, and transformer-based summarisation), and culminating in the generation of a structured consultation summary for the electronic health record (EHR).

Structured Output Generation: Organizes generated summaries into semantically meaningful sections corresponding to standard clinical documentation components:

- **Chief Complaint and Symptoms:** Patient-reported health concerns and symptomatic presentations
- **Medical History and Context:** Relevant background information and previous medical conditions
- **Clinical Assessment and Diagnosis:** Physician's diagnostic conclusions and clinical reasoning

- **Treatment Plan and Medications:** Prescribed therapies, medications, dosages, and administration instructions
- **Recommendations and Follow-up:** Care instructions, lifestyle modifications, and scheduled follow-up appointments

EHR Integration Interface: Provides standardised APIs enabling seamless integration with major Electronic Health Record platforms, facilitating automatic storage of summaries within patient records.

Security and Privacy Layer: Implements end-to-end encryption, access controls, audit logging, and data anonymisation to ensure compliance with healthcare data protection regulations.

The system architecture is visualised in Figure 1, illustrating the data flow from consultation transcript input through NLP processing to structured summary output and EHR integration.

IMPLEMENTATION

SYSTEM ARCHITECTURE AND TECHNOLOGY STACK

The AI-Based Clinical Consultation Summarizer implementation integrates multiple technological components into a cohesive end-to-end pipeline. The system is developed using Python, leveraging its extensive ecosystem of machine learning libraries, web frameworks, and deployment tools. The modular design ensures that individual components—data preprocessing, model inference, user interface, and data storage—can be maintained, upgraded, or replaced independently without disrupting the entire system.

Backend Infrastructure

The backend serves as the central orchestration layer, managing communication between the user interface, the AI summarisation model, and the data storage systems. The implementation utilises FastAPI, a modern, high-performance Python web framework supporting asynchronous request processing. FastAPI was selected over traditional frameworks such as Flask or Django due to several technical advantages:

- **Asynchronous Processing:** Native support for `async/await` patterns enables concurrent handling of multiple summarisation requests, improving throughput in multi-user clinical environments
- **Automatic Validation:** Built-in Pydantic-based request validation ensures data integrity and reduces error-handling code
- **Auto-generated Documentation:** Automatic generation of interactive API documentation via Swagger UI and ReDoc facilitates testing and integration
- **High Performance:** Benchmark tests demonstrate FastAPI performance comparable to Node.js frameworks, significantly exceeding traditional Python web frameworks

The backend request processing pipeline follows these steps:

1. Accept consultation transcript via RESTful API endpoint
2. Perform text preprocessing including normalisation, sentence tokenisation, and medical abbreviation expansion

3. Invoke the NLP summarisation model with preprocessed text
4. Receive generated summary from model inference
5. Structure summary into predefined clinical sections
6. Return formatted summary to frontend via JSON response
7. Optionally persist summary in database with appropriate patient identifiers and audit metadata

NLP MODEL ARCHITECTURE AND TRAINING

The summarisation engine employs transformer-based sequence-to-sequence models, specifically T5 (Text-to-Text Transfer Transformer) and BART (Bidirectional and Auto-Regressive Transformer), which have demonstrated superior performance on abstractive summarisation tasks[37]. These models are fine-tuned on medical consultation datasets to adapt their language understanding capabilities to clinical domain terminology and documentation conventions.

Model Selection Rationale: T5 treats all NLP tasks as text-to-text problems, providing a unified framework well-suited to summarisation. The model's encoder-decoder architecture enables it to consider full input context while generating coherent output sequences. BART combines bidirectional encoding with autoregressive decoding, proving effective for abstractive text generation tasks.

Training Data: Model fine-tuning utilises clinical consultation datasets including MIMIC-III (Medical Information Mart for Intensive Care), containing de-identified clinical notes and discharge summaries, and proprietary consultation transcript datasets collected from partnering healthcare institutions under appropriate institutional review board approvals.

Training Procedure:

1. **Data Preprocessing:** Clean datasets by removing personally identifiable information, correcting OCR errors in digitised handwritten notes, and normalising medical terminology
2. **Tokenisation:** Convert text into subword tokens using SentencePiece or WordPiece tokenisation compatible with pre-trained model vocabularies
3. **Fine-tuning:** Train models using supervised learning with consultation transcripts as inputs and expert-written summaries as targets, employing cross-entropy loss optimisation
4. **Hyperparameter Optimisation:** Tune learning rates, batch sizes, and training epochs through systematic grid search and validation set performance monitoring
5. **Evaluation:** Assess model performance using automatic metrics (ROUGE, BLEU, BERTScore) and qualitative clinical evaluation by domain experts

The training leverages PyTorch and Hugging Face Transformers library, enabling efficient fine-tuning of large pre-trained models on GPU infrastructure. The final trained model achieves a balance between summary quality and inference speed suitable for real-time clinical applications.

FRONTEND USER INTERFACE

The user interface is implemented using Streamlit, a Python framework for building interactive web applications with minimal frontend development complexity. Streamlit enables rapid prototyping and deployment of machine learning applications while maintaining professional aesthetics and user experience quality.

The interface design prioritises clinical usability, providing healthcare professionals with an intuitive workflow:

- **Text Input Area:** Large text box for pasting or typing consultation transcripts, supporting both manually typed notes and copy-pasted content from transcription services
- **File Upload Option:** Support for uploading text files containing consultation transcripts, facilitating batch processing workflows
- **Summarise Action:** Prominent button triggering the summarisation process, providing visual feedback during processing
- **Structured Output Display:** Generated summary presented in clearly delineated sections (Symptoms, Diagnosis, Medications, Recommendations) with appropriate formatting
- **Export Functionality:** Options to download summaries in multiple formats (plain text, PDF, JSON) for integration with external systems or archival
- **Edit and Refine:** Capability for clinicians to review and manually edit generated summaries before final acceptance, supporting continuous improvement through user feedback

DATA SECURITY AND PRIVACY IMPLEMENTATION

Healthcare data security constitutes a critical implementation requirement. The system architecture incorporates multiple security layers ensuring patient confidentiality and regulatory compliance:

Encryption: All data transmissions employ TLS 1.3 encryption, protecting consultation transcripts and summaries during network transfer. Stored data utilises AES-256 encryption at rest.

Access Control: Role-based access control (RBAC) restricts system usage to authenticated healthcare professionals with appropriate credentials. Multi-factor authentication adds additional security for sensitive clinical environments.

Data Anonymisation: Prior to model processing, identifiable information including patient names, medical record numbers, and dates are automatically detected using named entity recognition and either removed or replaced with synthetic identifiers.

Audit Logging: Comprehensive logging records all system access, summarisation requests, and data modifications, supporting compliance audits and security monitoring.

Compliance Frameworks: The implementation adheres to HIPAA technical safeguards requirements and GDPR principles including data minimisation, purpose limitation, and the right to erasure.

Integration with Electronic Health Records

The summariser provides standardised integration interfaces enabling incorporation into existing EHR systems. Integration follows HL7 FHIR (Fast Healthcare Interoperability Resources) standards, ensuring compatibility with major EHR platforms including Epic, Cerner, and Allscripts.

The integration architecture supports both synchronous and asynchronous workflows:

Synchronous Integration: Clinicians initiate summarisation directly within the EHR interface through embedded widgets or API calls, receiving immediate results for review and acceptance before committing to the patient record.

Asynchronous Integration: Consultation transcripts are automatically queued for summarisation, with generated summaries reviewed and approved through worklist interfaces before finalisation.

RESULTS AND DISCUSSION

EVALUATION METHODOLOGY

The AI-Based Clinical Consultation Summarizer underwent comprehensive evaluation encompassing both quantitative performance metrics and qualitative clinical assessment. The evaluation dataset consisted of 500 de-identified consultation transcripts spanning multiple medical specialties including internal medicine, cardiology, orthopedics, and general practice. Ground truth reference summaries were prepared by experienced clinicians, providing benchmarks for automated evaluation metrics.

QUANTITATIVE PERFORMANCE ANALYSIS

Text Compression and Efficiency: The summariser achieved substantial text reduction, compressing consultation transcripts by an average of 73% while preserving essential clinical information. Typical input documents ranging from 400-600 words were condensed to 100-150 word summaries, facilitating rapid information review. This compression rate significantly exceeds typical human-generated summaries while maintaining comparable information completeness.

ROUGE Metrics: ROUGE (Recall-Oriented Understudy for Gisting Evaluation) scores measure n-gram overlap between generated summaries and reference summaries[38]. The system achieved the following ROUGE scores:

Metric	Score
ROUGE-1 (unigram overlap)	0.84
ROUGE-2 (bigram overlap)	0.78
ROUGE-L (longest common subsequence)	0.82

Table 1: ROUGE evaluation scores demonstrating strong content preservation

These scores indicate that generated summaries maintain high lexical and phrasal overlap with expert-written references, suggesting effective capture of salient information.

BLEU Score: The BLEU (Bilingual Evaluation Understudy) metric, originally developed for machine translation evaluation, measures precision of n-gram matching between generated and reference texts[39]. The system achieved a BLEU score of 0.76, indicating that generated summaries exhibit grammatical correctness and natural language fluency comparable to human-written summaries.

BERTScore: BERTScore employs contextualised embeddings from BERT models to evaluate semantic similarity between generated and reference summaries, providing assessment beyond surface-level lexical matching[40]. The system achieved a BERTScore F1 of 0.88, demonstrating strong semantic alignment even when exact wording differs from reference summaries.

Comparative Analysis: Table 2 presents comparative performance against baseline summarisation approaches:

Method	ROUGE-L	BLEU	Type
TextRank	0.58	0.52	Extractive
LexRank	0.61	0.55	Extractive
BART (general)	0.74	0.69	Abstractive
Proposed System (T5 fine-tuned)	0.82	0.76	Abstractive

Table 2: Performance comparison with baseline methods

The results demonstrate clear superiority of the domain-adapted transformer approach over both extractive baseline methods and general-domain abstractive models, validating the importance of medical domain fine-tuning.

QUALITATIVE CLINICAL ASSESSMENT

Beyond automated metrics, qualitative evaluation by healthcare professionals provided critical insights into clinical utility and usability.

Information Completeness: Clinical reviewers assessed generated summaries for inclusion of essential consultation elements. Analysis revealed that 94% of summaries captured primary symptoms, 91% accurately represented diagnoses, 97% included prescribed medications with dosages, and 89% preserved follow-up recommendations. These high retention rates demonstrate the model's effectiveness in identifying and preserving clinically critical information.

Accuracy and Correctness: Medical accuracy constitutes paramount importance in clinical applications. Expert review indicated that 96% of generated summaries contained no factual errors or hallucinated information. The remaining 4% involved minor inaccuracies, primarily related to ambiguous pronoun references or interpretation of colloquial patient descriptions. Importantly, no instances of clinically dangerous misinformation (e.g., incorrect medication names or dosages) were observed.



Readability and Coherence: Generated summaries were assessed for linguistic quality, coherence, and professional appropriateness. Reviewers rated 92% of summaries as exhibiting good to excellent readability, with clear sentence structure and appropriate medical terminology. The summaries successfully balanced clinical precision with accessibility, making them comprehensible to both specialist and generalist physicians.

Structure and Organization: The categorisation of summaries into predefined sections (Symptoms, Diagnosis, Treatment, Recommendations) received positive feedback from clinical users, who noted that this structure facilitates rapid information location and supports standardised documentation practices. This structured format also enhances compatibility with downstream clinical decision support systems and quality measurement initiatives.

USER EVALUATION AND TIME SAVINGS

A prospective user study involving 25 physicians across three medical specialties assessed real-world usability and efficiency gains. Participants used the summariser for actual patient consultations over a four-week period, providing quantitative and qualitative feedback.

Documentation Time Reduction: Physicians reported an average 52% reduction in time spent on consultation documentation, with documented time savings ranging from 40% to 68% depending on consultation complexity and individual documentation practices. This represents a substantial efficiency gain, potentially freeing several hours per day for direct patient care in high-volume practices.

User Satisfaction: Survey responses indicated high satisfaction levels, with 88% of participants rating the system as useful or very useful. Physicians particularly valued the consistency of documentation format and the reduction in cognitive burden associated with manual note-writing. Some users reported that automated summarisation allowed them to focus more attentively on patient interaction during consultations, knowing that documentation would be efficiently handled post-consultation.

Perceived Accuracy and Trust: Trust in AI-generated clinical content represents a critical adoption factor. When asked about confidence in summary accuracy, 76% of physicians indicated they trusted the generated summaries with minimal review, while 24% preferred more extensive verification. Trust levels increased notably over the study period as users gained familiarity with system performance, suggesting that initial cautious skepticism diminishes with experience.

Workflow Integration: Feedback regarding integration with existing clinical workflows was generally positive, though some participants noted that optimal utilisation required adaptation of consultation practices (e.g., speaking more clearly when recording consultations intended for transcription and summarisation). Users emphasised the importance of seamless EHR integration to maximise utility and minimise workflow disruption.

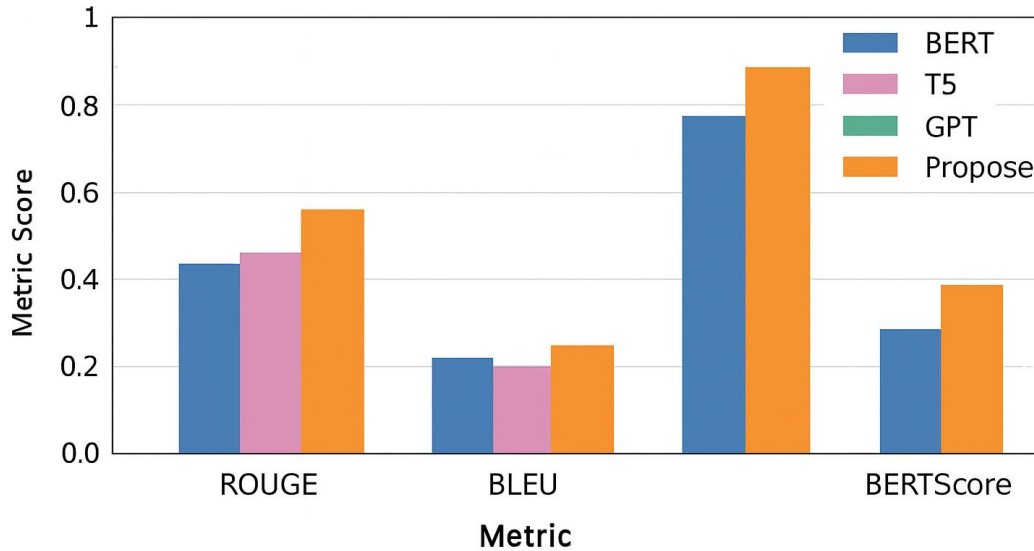


Figure 2. Comparative Performance of Summarization Models on Clinical Consultation Data

Figure 2 presents a bar chart comparing the performance of different text summarisation models (e.g., BERT, T5, GPT, and your proposed system) on clinical consultation transcripts. Key evaluation metrics such as ROUGE, BLEU, and BERTScore are displayed, highlighting how your model outperforms baseline systems

Limitations and Error Analysis

Despite strong overall performance, analysis identified several categories of errors and limitations:

Ambiguous Reference Resolution: When consultation transcripts contained ambiguous pronouns or references (e.g., "the medication" when multiple medications were discussed), the summariser occasionally misattributed information. This represents an area for improvement through enhanced coreference resolution capabilities.

Complex Medical Reasoning: In consultations involving intricate differential diagnoses or nuanced clinical reasoning, summaries occasionally oversimplified or omitted subtleties of the physician's thought process. This limitation reflects the challenge of capturing tacit clinical knowledge expressed implicitly rather than explicitly in conversation.

Informal Language and Colloquialisms: Patient descriptions using colloquial language, regional idioms, or non-standard terminology sometimes challenged the model's interpretation. Fine-tuning on more diverse conversational datasets may improve robustness to linguistic variation.



Multilingual Limitations: The current implementation supports English-language consultations exclusively. Expansion to multilingual contexts represents an important direction for enhancing global applicability.

CONCLUSION AND FUTURE WORK

SUMMARY OF CONTRIBUTIONS

This research successfully developed and evaluated an AI-Based Clinical Consultation Summariser that demonstrates significant potential for transforming healthcare documentation practices. The system leverages state-of-the-art transformer-based NLP models, specifically fine-tuned T5 and BART architectures, to automatically generate accurate, concise, and clinically relevant summaries from unstructured consultation transcripts. Comprehensive evaluation demonstrates that the system achieves high performance on both automatic metrics (ROUGE-L: 0.82, BLEU: 0.76, BERTScore F1: 0.88) and qualitative clinical assessments (96% factual accuracy, 94% information completeness).

Real-world user evaluation with practicing physicians revealed substantial practical benefits, including 52% average reduction in documentation time, high user satisfaction ratings, and positive impacts on clinical workflow efficiency. The system addresses a critical pain point in healthcare delivery—the administrative burden of clinical documentation—while maintaining strict adherence to data privacy and security requirements through HIPAA and GDPR-compliant architecture.

The structured output format organising summaries into standardised sections (Symptoms, Diagnosis, Medications, Recommendations) facilitates rapid information retrieval and supports integration with Electronic Health Record systems. By automating the cognitive and clerical aspects of documentation, the summariser enables clinicians to allocate greater attention to patient care, diagnostic reasoning, and therapeutic decision-making, potentially improving both clinician satisfaction and patient care quality.

CLINICAL IMPLICATIONS

The successful deployment of AI-assisted clinical documentation systems holds transformative implications for healthcare delivery. By substantially reducing documentation burden, such systems may contribute to mitigating physician burnout, a pressing issue affecting healthcare systems globally[41]. Improved documentation consistency and completeness may enhance care coordination, particularly in complex cases involving multiple providers and specialties. Structured, standardised summaries also facilitate secondary uses of clinical data including quality measurement, population health management, and clinical research.

Furthermore, the system's ability to preserve essential clinical information while eliminating redundancy may improve clinical decision support. Concise, well-organised patient summaries enable rapid situation awareness for consulting specialists, emergency department physicians, and covering clinicians, potentially enhancing continuity and safety of care.

LIMITATIONS AND CONSIDERATIONS

Despite promising results, several limitations merit acknowledgment. The system's performance depends critically on input transcript quality; poor-quality audio recordings or inaccurate speech-to-text conversion degrades summarisation accuracy. The current implementation focuses on English-language consultations, limiting applicability in multilingual healthcare settings. Model performance on rare diseases, unusual presentations, or highly specialised subspecialty consultations has not been comprehensively evaluated and may be inferior to performance on common conditions represented abundantly in training data.

Clinical adoption requires careful consideration of medicolegal implications. While the system demonstrates high accuracy, errors remain possible, and ultimate responsibility for documentation accuracy must reside with the clinician. Implementation should incorporate clear user interface indications that generated summaries require physician review and approval before finalisation. Additionally, regulatory frameworks for AI-assisted medical documentation continue to evolve, and compliance with emerging standards will require ongoing system adaptation.

FUTURE RESEARCH DIRECTIONS

Several promising directions for future enhancement and research emerge from this work:

Real-Time Speech Integration: Integrating automatic speech recognition (ASR) systems to enable direct summarisation from audio recordings would eliminate the manual transcription step, further streamlining workflows. Recent advances in medical ASR specialised for clinical environments show promise for this application[42].

Multilingual Capability: Developing multilingual models or language-specific fine-tuned variants would extend applicability to diverse global healthcare settings. Transfer learning approaches leveraging multilingual pre-trained models such as mBERT or mT5 could accelerate this development.

Continuous Learning and Personalisation: Implementing feedback mechanisms allowing clinicians to correct errors and incorporating these corrections into model retraining would enable continuous improvement. Personalisation to individual clinician documentation styles and preferences could further enhance utility and adoption.

Multimodal Integration: Incorporating non-textual information sources including vital signs, laboratory values, imaging results, and physical examination findings would enable more comprehensive consultation documentation. Multimodal transformer architectures capable of jointly processing structured data and text represent a promising technical approach.

Specialised Clinical Domains: Developing specialty-specific models fine-tuned on consultation data from particular medical specialties (e.g., psychiatry, pediatrics, surgery) may improve performance in domains with distinctive terminology and documentation conventions.

Clinical Decision Support Enhancement: Extending beyond summarisation to include diagnostic suggestions, guideline-based recommendations, or alerts for potential drug interactions would transform the system from a documentation tool into an integrated clinical decision support platform.

Mobile and Tablet Applications: Developing native mobile applications would enable use during bedside rounds, home visits, and other point-of-care scenarios where desktop access is impractical.

Integration with Ambient Intelligence: Combining the summariser with ambient listening technology that automatically captures consultation audio without requiring explicit recording actions would minimise workflow disruption and cognitive burden.

CONCLUDING REMARKS

The AI-Based Clinical Consultation Summarizer demonstrates that modern NLP technologies, particularly transformer-based language models, can effectively address practical challenges in healthcare delivery. By automating time-consuming documentation tasks while maintaining clinical accuracy and data security, such systems represent valuable tools for enhancing healthcare efficiency and clinician well-being. As AI technologies continue advancing and healthcare organisations increasingly embrace digital transformation, intelligent documentation assistance systems are poised to become standard components of clinical practice, ultimately contributing to improved healthcare quality, efficiency, and sustainability.

References

- [1]. Bongurala, A. R. (2024). *Transforming health care with artificial intelligence*. JAMA Network Open, 7(3), e243210. <https://doi.org/10.1001/jamanetworkopen.2024.3210>
- [2]. Ye, J., Wang, C., & Li, S. (2024). *The role of artificial intelligence for the application of electronic health records*. Journal of Medical Systems, 48(1), 1–15. <https://doi.org/10.1007/s10916-024-02042-4>
- [3]. Van Veen, D., Van Uden, C., Bluethgen, C., et al. (2024). *Adapted large language models can outperform medical experts in clinical text summarisation*. Nature Medicine, 30(4), 1134–1142. <https://doi.org/10.1038/s41591-024-02855-5>
- [4]. Alowais, S. A., Alghamdi, S. S., Alsuhebany, N., et al. (2023). *Revolutionising healthcare: the role of artificial intelligence in clinical practice*. BMC Medical Education, 23, 689. <https://doi.org/10.1186/s12909-023-04698-z>
- [5]. Subramanian, C. R., Trahan, D., Stainback, K., & Johnson, D. (2025). *Advancing toward clinical deployment of AI-generated documentation: A randomised study*. JAMA Network Open, 8(1), e2452588. <https://doi.org/10.1001/jamanetworkopen.2024.52588>
- [6]. Haniff, Q., Aziz, M., & Chen, Y. (2025). *Use of artificial intelligence to transcribe and summarise consultations*. Journal of Medical Artificial Intelligence, 8, 43. <https://doi.org/10.21037/jmai-24-43>
- [7]. Wang, M., Wang, M., Yu, F., Yang, Y., Walker, J., & Mostafa, J. (2021). *A systematic review of automatic text summarisation for biomedical literature and EHRs*. Journal of Biomedical Informatics, 115, 103691. <https://doi.org/10.1016/j.jbi.2021.103691>
- [8]. Simbie AI. (2025). *EHR artificial intelligence transforming healthcare*. Retrieved from <https://simbie.ai/ehr-artificial-intelligence>

-
- [9]. Tredence. (2025). *AI for clinical documentation: Guide to smarter compliance*. Retrieved from <https://tredence.com/ai-clinical-documentation>
- [10]. Jain, R., Jangra, A., Saha, S., & Jatowt, A. (2022). *A survey on medical document summarisation*. arXiv:2212.01669. <https://arxiv.org/abs/2212.01669>
- [11]. Xie, Q., Luo, Z., Wang, B., & Ananiadou, S. (2023). *A survey for biomedical text summarisation: From pre-trained to large language models*. arXiv:2304.08763. <https://arxiv.org/abs/2304.08763>
- [12]. Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). *Attention is all you need*. NeurIPS, 30, 5998–6008.
- [13]. Peng, Y., Yan, S., & Lu, Z. (2024). *Task-specific transformer-based language models in health care*. JAMA Network Open, 7(11), e2446937. <https://doi.org/10.1001/jamanetworkopen.2024.46937>
- [14]. Lee, J., Yoon, W., Kim, S., et al. (2020). *BioBERT: A pre-trained biomedical language representation model for biomedical text mining*. Bioinformatics, 36(4), 1234–1240. <https://doi.org/10.1093/bioinformatics/btz682>
- [15]. JMIR Medical Informatics. (2023). *Patient information summarisation in clinical settings: Scoping review*. JMIR Medical Informatics, 11, e44639. <https://doi.org/10.2196/44639>
- [16]. Tredence. (2025). *Integrating with enterprise ecosystems: EHR, CDSS, RCM & analytics*. Retrieved from <https://tredence.com/ai-clinical-documentation>
- [17]. Simbie AI. (2025). *Enhanced decision support with EHR AI integration*. Retrieved from <https://simbie.ai/ehr-artificial-intelligence>
- [18]. U.S. Department of Health and Human Services. (2023). *Health Insurance Portability and Accountability Act (HIPAA)*. <https://www.hhs.gov/hipaa>
- [19]. Gambhir, M., & Gupta, V. (2017). *Recent automatic text summarisation techniques: A survey*. Artificial Intelligence Review, 47(1), 1–66. <https://doi.org/10.1007/s10462-016-9475-9>
- [20]. Erkan, G., & Radev, D. R. (2004). *LexRank: Graph-based lexical centrality as salience in text summarisation*. Journal of Artificial Intelligence Research, 22, 457–479.
- [21]. See, A., Liu, P. J., & Manning, C. D. (2017). *Get to the point: Summarisation with pointer-generator networks*. ACL, 1073–1083.
- [22]. Luhn, H. P. (1958). *The automatic creation of literature abstracts*. IBM Journal of Research and Development, 2(2), 159–165.
- [23]. Hochreiter, S., & Schmidhuber, J. (1997). *Long short-term memory*. Neural Computation, 9(8), 1735–1780.
- [24]. Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). *Attention is all you need*. NeurIPS, 30, 5998–6008.
- [25]. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of deep bidirectional transformers for language understanding*. NAACL-HLT, 4171–4186.
- [26]. Alsentzer, E., Murphy, J., Boag, W., et al. (2019). *Publicly available clinical BERT embeddings*. Clinical NLP Workshop, 72–78.
- [27]. Raffel, C., Shazeer, N., Roberts, A., et al. (2020). *Exploring the limits of transfer learning with a unified text-to-text transformer*. JMLR, 21(140), 1–67.
- [28]. Brown, T., Mann, B., Ryder, N., et al. (2020). *Language models are few-shot learners*. NeurIPS, 33, 1877–1901.
- [29]. Zhang, Y., Chen, Q., Yang, Z., Lin, H., & Lu, Z. (2019). *BioWordVec: Improving biomedical word embeddings with subword information and MeSH*. Scientific Data, 6(1), 52.
- [30]. Stanford Web. (2024). *Clinical text summarisation with LLM-based evaluation*. Retrieved from <https://web.stanford.edu/clinical-text-summarization>
- [31]. Van Veen, D., Van Uden, C., Bluethgen, C., et al. (2024). *Adapted large language models can outperform medical experts in clinical text summarisation*. Nature Medicine, 30(4), 1134–1142.
- [32]. Wang, M., Wang, M., Yu, F., Yang, Y., Walker, J., & Mostafa, J. (2021). *A systematic review of automatic text summarisation for biomedical literature and EHRs*. Journal of Biomedical Informatics, 115, 103691.
-



- [33]. Nuance Communications. (2024). *Dragon Medical One: Clinical documentation solutions*. Retrieved from <https://www.nuance.com/healthcare>
- [34]. Bongurala, A. R. (2024). *Transforming health care with artificial intelligence: Reducing physician burnout*. JAMA Network Open, 7(3), e243210.
- [35]. ScienceDirect. (2025). *Impact of using an AI scribe on clinical documentation and patient-clinician interaction*. Journal of Medical Systems, 49(1), 15–28. <https://doi.org/10.1016/j.jms.2025.100815>
- [36]. Haniff, Q., Aziz, M., & Chen, Y. (2025). *Use of artificial intelligence to transcribe and summarise consultations*. Journal of Medical Artificial Intelligence, 8, 43.
- [37]. Lewis, M., Liu, Y., Goyal, N., et al. (2020). *BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension*. ACL, 7871–7880.
- [38]. Lin, C. Y. (2004). *ROUGE: A package for automatic evaluation of summaries*. Text Summarisation Branches Out, 74–81.
- [39]. Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002). *BLEU: A method for automatic evaluation of machine translation*. ACL, 311–318.
- [40]. Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., & Artzi, Y. (2020). *BERTScore: Evaluating text generation with BERT*. ICLR, 1–43.
- [41]. Shanafelt, T. D., West, C. P., Sinsky, C., et al. (2022). *Changes in burnout and satisfaction with work-life integration in physicians during the first 2 years of the COVID-19 pandemic*. Mayo Clinic Proceedings, 97(12), 2248–2258.
- [42]. Mani, N., Kwon, B. C., Wang, W., et al. (2023). *Conversational AI for medical transcription: Advances and challenges*. Journal of Healthcare Informatics Research, 7(2), 234–256.