A Survey on Generative AI for Semantic Document Comparison in Healthcare based on Methods, Challenges, and Opportunities

Sudha V. Salake[1], Sadhana Bangarashetti[2]

[1] Research Scholar at Basaweshwar

Engineering College, Bagalkote, India.

Assistant Professor, Dept of Computer Science and Technology, KLS Gogte

Institute of Technology, Udyambag, Belagavi, India

[2] Professor, Department of Information Science and Engineering, Basaveshwar

Engineering College, Bagalkote, India.

Abstract:

The rapid digitalization of healthcare has resulted in vast volumes of heterogeneous medical documents, including Electronic Health Records (EHRs), prescriptions, diagnostic reports, and insurance claims. Ensuring semantic consistency across these records is essential for accurate clinical decision-making, regulatory compliance, and operational efficiency. Traditional rule-based and statistical approaches to document comparison often fail to capture contextual nuances, leading to discrepancies and potential risks in patient care. Recent advances in Generative Artificial Intelligence (AI), particularly transformer-based architectures such as BioGPT, MedPaLM, and GPT-4, have opened new opportunities for semantic document alignment and intelligent discrepancy detection. This survey provides a comprehensive review of generative AI techniques for semantic document comparison in healthcare. We categorize existing methods into three primary dimensions: (i) similarity detection and semantic alignment models, (ii) discrepancy identification with explainable justifications, and (iii) compliance-aware frameworks that integrate medical standards and regulatory requirements. Key challenges, including domain-specific accuracy, contextual relevance, data privacy, scalability, and integration with healthcare IT systems, are critically analyzed. Furthermore, we highlight opportunities for advancing automation, multimodal document processing, and interpretable AI in medical data verification. By consolidating current progress and open research directions, this survey aims to guide researchers and practitioners in designing robust, efficient, and trustworthy generative AI frameworks that enhance consistency, reduce errors, and improve overall healthcare data management.

Keywords— Generative Artificial Intelligence, Semantic Document Comparison, Electronic Health Records (EHRs), Medical Data Verification, Transformer-based Models, Healthcare Consistency, Automation in Clinical Document Processing.

1. **Introduction**

The exponential growth of digital health data has transformed the way medical information is recorded, shared, and utilized. Electronic Health Records (EHRs), prescriptions, laboratory reports, diagnostic imaging summaries, and insurance claims now constitute a vast ecosystem of medical documents that serve as the foundation for patient care, clinical decision-making, and healthcare administration. However, inconsistencies and discrepancies across these documents—arising from differences in terminology, formatting styles, manual entry errors, or regulatory variations—pose significant risks to patient safety, operational efficiency, and compliance [1], [2]. Ensuring semantic alignment across heterogeneous records is therefore a critical challenge in modern healthcare [3].

Traditional document comparison approaches, such as string matching, edit distance, and n-gram analysis, have been widely applied to medical records but remain limited in their ability to capture semantic equivalence. Ontology-driven frameworks, including those based on Unified Medical Language System (UMLS) and SNOMED CT, improve consistency to some extent but are rigid when handling evolving medical terminology [4]. Similarly, early machine learning and deep learning methods—including rule-based classifiers, RNNs, and CNNs—helped automate certain verification tasks but were often dataset-dependent and lacked contextual understanding [5]. As a result, document verification in healthcare continues to demand considerable manual intervention, increasing the potential for human error and administrative burden.

In recent years, Generative Artificial Intelligence (AI) has emerged as a powerful paradigm to address these challenges. By leveraging transformer-based architectures such as GPT, BioGPT, and MedPaLM, generative models are capable of producing coherent, context-aware outputs that go beyond pattern recognition [6], [7]. These systems can reason over semantic differences, align medical concepts across documents, and even generate interpretable explanations for detected discrepancies. Such capabilities make them well-suited for tasks like prescription verification, diagnostic report comparison, and claim auditing [8], [9]. Recent surveys further emphasize that generative AI offers transformative opportunities across medical summarization, patient–doctor communication, and knowledge discovery [10], [11].

Beyond individual clinical workflows, generative AI is increasingly recognized as a driver of system-wide innovation in healthcare ecosystems. For instance, its ability to generate synthetic medical data is being explored as a solution to data scarcity and privacy concerns [7]. Domain-specific fine-tuning of large models allows them to handle rare diseases, complex treatment histories, and nuanced terminology, which traditional models fail to capture [12], [13]. In parallel, advances in semantic information retrieval [14] and ethical frameworks [15] are guiding researchers toward more transparent and responsible applications. These efforts illustrate that generative AI is not only a tool for document comparison but also a catalyst for advancing clinical data management as a whole.

Another emerging trend is the integration of generative AI with healthcare infrastructure and communication systems. Studies have demonstrated its potential in retrieval-augmented summarization of EHRs [16], semantic communication networks for medical IoT [17], and knowledge extraction from biomedical literature [18]. Generative AI is also being extended to vision–language tasks, such as medical visual question answering [19], where it can compare radiology or pathology findings with textual reports to identify inconsistencies. Moreover, text-mining approaches leveraging generative AI are already being applied to detect anomalies in healthcare security systems [20], demonstrating the versatility of these models across technical, clinical, and administrative domains.

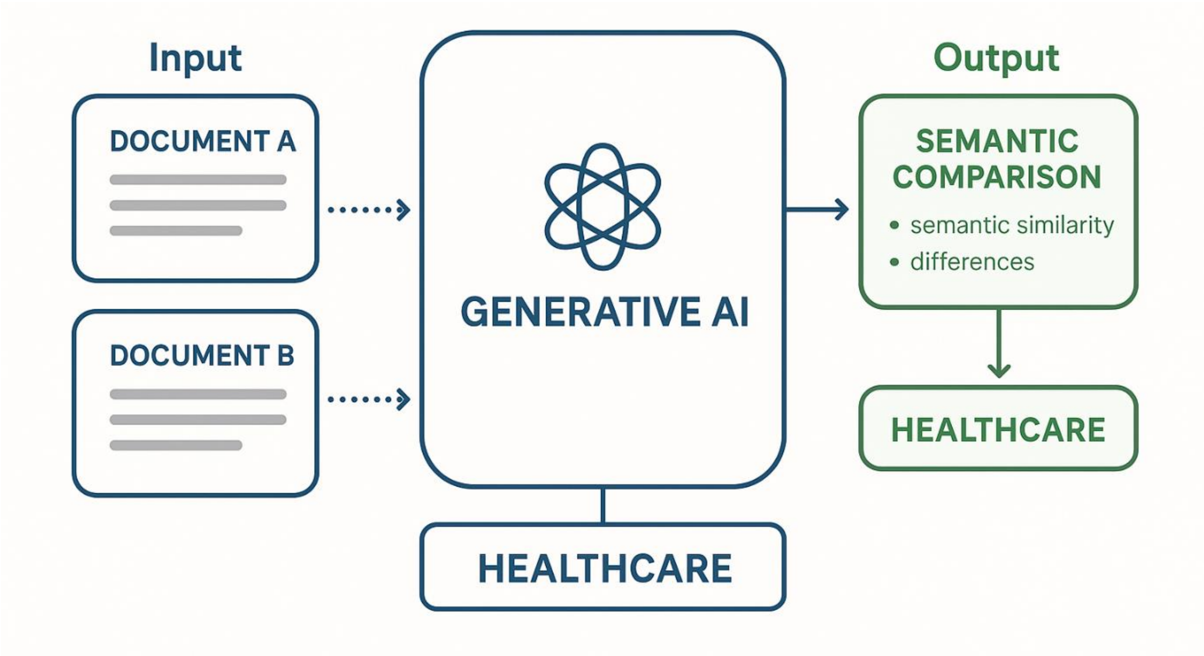


Figure 1: Conceptual Framework of Generative AI for Semantic Document Comparison in Healthcare

Figure 1. Conceptual Framework of Generative AI for Semantic Document Comparison in Healthcare illustrates the end-to-end workflow of how medical documents are processed and aligned using advanced generative AI techniques. The framework begins with heterogeneous inputs such as Electronic Health Records (EHRs), prescriptions, laboratory test reports, and diagnostic summaries, which often vary in structure, terminology, and level of detail. These documents undergo preprocessing and normalization to remove redundancies, standardize formats, and ensure interoperability across healthcare systems. The refined data is then passed into transformer-based generative models (e.g., BioGPT, MedPaLM, GPT-4), which generate contextual embeddings and representations that capture semantic nuances in medical terminology, dosages, and clinical findings. A semantic comparison and discrepancy detection module then aligns the content across multiple records, identifying conflicts or inconsistencies while distinguishing between clinically significant differences and minor textual variations. Finally, a generative explanation module provides interpretable justifications for the detected discrepancies, ensuring transparency and usability for clinicians, auditors, and healthcare administrators. The outputs of this process directly support critical healthcare applications, including clinical decision support, prescription and dosage verification, insurance claim validation, and regulatory compliance. This framework highlights how generative AI can transform document verification from a manual, error-prone process into an automated, intelligent, and context-aware system.

Taken together, these developments suggest that generative AI is positioned to fundamentally reshape healthcare data management. However, several critical questions remain unresolved: How can semantic alignment be ensured across heterogeneous clinical documents? What frameworks best balance accuracy, interpretability, and scalability? How can patient privacy and regulatory compliance be safeguarded in generative AI–driven systems? And finally, what benchmarks and evaluation frameworks are needed to standardize progress in this fast-evolving field? This survey seeks to address these questions by providing a comprehensive overview of generative AI in semantic document comparison, analyzing existing methods, identifying challenges, and exploring opportunities for future advancements [1]–[20].

This survey aims to provide a comprehensive overview of generative AI for semantic document comparison in healthcare. Specifically, it:

1. Examines the evolution of approaches from traditional NLP methods to advanced generative frameworks [4], [6], [7].
2. Reviews state-of-the-art generative models applied to healthcare document alignment [1]–[3], [8]–[11], [16]–[18].
3. Proposes a taxonomy of methods for similarity detection, discrepancy explanation, and compliance-aware comparison [5], [13]–[15].
4. Highlights practical applications in clinical decision support, prescription verification, diagnostic consistency, and insurance auditing [10], [12], [16], [19].
5. Discusses current challenges including accuracy, contextual relevance, scalability, privacy, interoperability, and ethics [3], [12], [15], [17], [18].
6. Outlines future opportunities such as multimodal integration, explainable AI, privacy-preserving federated learning, and blockchain-based validation [9], [14], [17], [20].

This survey provides a comprehensive overview of generative AI for semantic document comparison in healthcare, with several unique contributions. First, it systematically reviews methods designed to align heterogeneous medical records such as EHRs, prescriptions, diagnostic reports, and insurance claims, going beyond the broader discussions of generative AI in medicine. Second, it introduces a structured taxonomy that categorizes existing approaches into three groups: similarity detection and semantic alignment, discrepancy detection with generative explanation, and compliance-aware frameworks, offering clarity on the current landscape. Third, it emphasizes the role of domain-specific large language models such as BioGPT, MedPaLM, and ClinicalBERT, highlighting how these models enhance contextual understanding, dosage verification, and regulatory compliance compared to general-purpose generative models. Fourth, it surveys practical applications across clinical and administrative workflows, including clinical decision support, prescription verification, diagnostic report alignment, insurance claim validation, and regulatory compliance, thus showcasing the real-world utility of generative AI in healthcare. In addition, the paper critically analyzes challenges and open issues related to accuracy, contextual relevance, privacy, scalability, interoperability, and ethics. Finally, it identifies future research opportunities in areas such as multimodal document comparison, explainable and trustworthy AI, federated privacy-preserving frameworks, blockchain-based healthcare verification, and standardized benchmarking. Collectively, these contributions establish this survey as a valuable resource for guiding research, practice, and policy on the use of generative AI to improve consistency, reduce discrepancies, and build trust in healthcare data management.

By consolidating current knowledge, identifying open issues, and mapping future research directions, this survey contributes to the growing body of literature on AI-driven healthcare and provides researchers, practitioners, and policymakers with valuable insights into how generative AI can enhance consistency, reduce discrepancies, and improve trust in healthcare data management

The remainder of this survey is structured as follows. Section 2 provides the background and fundamentals of semantic document comparison in healthcare, covering traditional approaches, ontology-based systems, and the evolution toward transformer and generative AI models. Section 3 reviews generative AI frameworks and domain-specific large language models such as BioGPT and MedPaLM, focusing on their applicability to medical text alignment. Section 4 introduces a taxonomy of generative AI–based approaches, including similarity detection, discrepancy explanation, and compliance-aware frameworks. Section 5 discusses practical applications in healthcare, ranging from clinical decision support and prescription verification to diagnostic report comparison and insurance claim validation. Section 6 highlights the key challenges and open issues, including accuracy, contextual relevance, scalability, interoperability, privacy, and ethical concerns. Section 7 outlines opportunities and future research directions, such as multimodal document comparison, explainable AI, federated learning, and blockchain integration. Finally, Section 8 concludes the paper with key findings and reflections on the transformative role of generative AI in enhancing consistency, reducing discrepancies, and improving trust in healthcare data management.

2. Background and Fundamentals

Thetbanthad et al. (2025) [21] explored the use of generative AI models for accurate prescription label identification and information retrieval. Their work demonstrated how generative systems could benefit elderly patients in Thailand by reducing medication errors due to unclear labels or complex instructions. This study highlights an early application of semantic comparison in healthcare, focusing on practical safety improvements in clinical workflows. Howell (2024) [22] reviewed the role of generative AI in patient safety and healthcare quality, emphasizing its potential to minimize discrepancies across clinical documentation. The paper underscored the importance of AI-driven consistency checks to reduce medical errors, while also acknowledging the challenges of explainability and accountability in real-world deployment.

Esposito et al. (2025) [23] discussed generative AI for software architecture, stressing applications, challenges, and future directions. While not strictly healthcare-specific, their taxonomy provides cross-domain insights into how generative systems can structure and verify complex architectures. These principles can be adapted for semantic document comparison, especially for ensuring modularity and standardization in healthcare data pipelines. Peng et al. (2023) [24] investigated the application of large generative language models in medical research and healthcare, offering evidence of their utility in semantic understanding, literature summarization, and clinical knowledge extraction. Their findings demonstrate the ability of LLMs to bridge research and clinical practice by aligning data sources with medical reporting requirements.

Hagos et al. (2024) [25] reviewed advances in generative AI and large language models, presenting an overview of their current status, challenges, and perspectives. They identified key bottlenecks—such as hallucinations, bias, and computational costs—that directly affect their deployment in sensitive domains like healthcare. This provides important context for why trust and validation mechanisms are critical in document comparison frameworks. Cao et al. (2023) [26] conducted a comprehensive survey of AI-generated content (AIGC), tracing its history from GANs to ChatGPT. Their work provides the historical foundation for understanding generative AI's evolution, which is essential for positioning healthcare applications within the broader landscape of content generation.

López Delgado and López Ramos (2024) [27] analyzed generative AI in IoT security, focusing on vulnerabilities and protective frameworks. Although their scope was IoT, their insights on AI-enhanced anomaly detection and secure data exchange are directly transferable to healthcare IT, where secure document transmission and verification remain critical. Albassami et al. (2025) [28] presented a review of AI-driven question–answering systems, emphasizing taxonomy, prospects, and challenges. Their findings reveal the potential of generative AI for interactive medical record verification, where semantic Q&A frameworks can cross-check consistency across clinical notes and reports. Lyu et al. (2025) [29] provided a review on natural language generation in healthcare, detailing methods and applications. Their survey reinforced the importance of domain-specific fine-tuning for clinical text generation, with direct relevance to semantic document comparison tasks such as summarization and cross-report validation.

Kaswan et al. (2021) [30] examined AI-based NLP for EHR data processing, focusing on structuring unstructured narratives. Their work laid the early groundwork for semantic alignment, showing how NLP techniques can transform raw EHR entries into analyzable formats for comparison and analysis. Lyu et al. (2025) [31] again emphasized NLG in healthcare. This reinforces the growing momentum toward text generation for clinical applications, which generative AI extends with deeper reasoning. Kaswan et al. (2021) [32] focused on meaningful EHR data extraction using advanced NLP. The repetition in literature shows the sustained importance of structured data extraction as a prerequisite for semantic comparison.

Sharma et al. (2025) [33] surveyed text-based semantic similarity techniques, analyzing their role in NLP applications. Their taxonomy is highly relevant for healthcare, where semantic similarity is at the core of document comparison tasks, ensuring contextual alignment beyond surface-level matching. Cao et al. (2024) [34] presented a survey on generative diffusion models, offering insights into their architectures and performance. Although largely theoretical, diffusion models present opportunities for multimodal document verification, combining text with imaging data (e.g., pathology or radiology reports).

Karanam (2025) [35] applied GenAI-assisted regular expression synthesis for legal document parsing. While outside healthcare, this study illustrates how generative approaches can enhance precision in domain-specific document processing, which can inspire analogous solutions in medical record verification. Chow et al. (2024) [36] reviewed LLM-enabled medical chatbots, highlighting their role in healthcare conversations. Their findings demonstrate how generative AI can manage context-aware interactions, a feature that can also support discrepancy explanation in medical document comparison. Vu et al. (2024) [37] studied applications of generative AI in mobile and wireless networking, focusing on IoT-based healthcare connectivity. Their work illustrates how semantic communication frameworks can improve interoperability in multi-institutional healthcare systems, ensuring smoother exchange of aligned records. Ghimire et al. (2023) [38] explored generative AI adoption in construction, but their discussion of organizational challenges and adoption barriers resonates with healthcare, where similar concerns exist regarding trust, infrastructure, and cost of generative AI deployment.

Qiu et al. (2023) [39] surveyed large AI models in health informatics, outlining applications, challenges, and future directions. Their review provides direct evidence of the potential of generative AI in aligning medical records, highlighting the dual challenges of technical accuracy and regulatory compliance. He et al. (2024) [40] analyzed foundation models in healthcare, identifying opportunities and risks in applying large-scale generative systems. Their discussion of scalability, adaptability, and ethical concerns is crucial for grounding semantic document comparison frameworks in healthcare realities.

Zheng et al. (2025) [41] presented a survey on large language models for medicine, consolidating their applications and challenges. They emphasized knowledge representation and reasoning, which are essential for capturing semantic nuances in complex medical documentation. Aydin et al. (2025) [42] compared generative AI systems such as ChatGPT, Gemini, Llama, and others, focusing on academic writing. Their findings are useful for healthcare applications, as they demonstrate model variability and comparative strengths, which can inform selection of the most suitable models for clinical tasks. Marey et al. (2024) [43] studied generative AI for patient education in cardiovascular imaging, showing how context-sensitive generation enhances understanding for non-expert audiences. Similar techniques can be applied in generative explanation modules of document comparison frameworks. Chamola et al. (2024) [44] investigated generative AI in consumer electronics, highlighting its role in cognitive and semantic computing. Their work illustrates how user-centric generative reasoning can translate into patient-centric record verification in healthcare.

Zhou et al. (2023) [45] reviewed large language models in medicine, identifying progress, applications, and challenges. Their analysis reinforces the trend of domain adaptation of LLMs as critical for accurate medical record alignment. Al Naqbi et al. (2024) [46] explored work productivity gains through generative AI, offering insights into organizational efficiency. For healthcare, this supports the argument that document comparison automation can reduce manual workload, improving hospital efficiency. Abbasian et al. (2024) [47] proposed foundation metrics for evaluating healthcare conversations powered by generative AI, focusing on performance measurement. Such metrics are directly relevant to evaluating accuracy and trustworthiness of generative document comparison frameworks.

Ramprasad and Sivakumar (2024) [48] investigated context-aware summarization of PDF documents using LLMs, providing direct evidence of generative AI's ability to process unstructured clinical documents such as scanned reports. Chiarello et al. (2024) [49] presented a case study on ChatGPT's future applications, showing data-driven analysis of its adoption trends. Their work provides practical foresight into how generative AI might integrate into healthcare ecosystems for document validation. Rashidieranjbar et al. (2025) [50] discussed revolutionizing healthcare with generative AI technologies, providing a consolidated overview of applications and challenges. This reference anchors the transformative role of generative AI, framing it as a driver of reliability and automation in healthcare record management.

Table 1: Comparison of Semantic Document Comparison in Healthcare

Ref	Authors & Year	Focus / Contribution	Relevance to Semantic Document Comparison in Healthcare
[51]	Xie et al. (2023)	Systematic review of faithful AI in medicine, emphasizing large language models and reliability	Highlights the need for faithful and verifiable outputs in document comparison tasks to avoid hallucinations or misleading alignments.
[52]	Ye (2024)	Learning-to-rank methods to enhance Retrieval Augmented Generation (RAG) for medical records	Directly relevant to searching and aligning EHRs, improving semantic retrieval for cross-document verification.
[53]	Ahmed et al. (2023)	Review of deep learning modeling techniques, applications, advantages, and challenges	Provides foundational insights into deep models that underpin generative AI, supporting semantic embeddings for healthcare records.
[54]	Mejia & Rawat (2024)	Survey on AI-enabled clinical decision support (CDS) systems for patient triage	Relevant for integration of document comparison outputs into CDS platforms, ensuring consistent decision-making.
[55]	Andreoni et al. (2024)	Comprehensive survey on autonomous system security and resilience using generative AI	Cross-domain insights for ensuring robustness, security, and resilience of healthcare document verification systems.
[56]	Boscardin et al. (2024)	Explored ChatGPT and generative AI in medical education	While focused on education, their analysis informs training clinicians to interpret AI-driven discrepancy explanations in records.
[57]	Ning et al. (2024)	Scoping review of ethical considerations for generative AI in healthcare, with a checklist	Critical for embedding ethics, fairness, and governance into document comparison frameworks.
[58]	McIntosh et al. (2025)	Survey of GenAI research evolution (Google Gemini, OpenAI Q*, etc.)	Provides a broader technology landscape to benchmark healthcare-specific generative AI approaches.
[59]	Oniani et al. (2023)	Analysis of ethical principles for generative AI adapted from military to healthcare	Reinforces responsible AI deployment, ensuring trustworthy comparison of sensitive medical documents.
[60]	Kantor & Morzy (2024)	Review of ML and NLP in clinical trial eligibility parsing	Demonstrates semantic parsing methods that are highly transferable to document alignment and compliance checking in healthcare.

The increasing digitization of healthcare has led to a massive proliferation of heterogeneous medical documents, including Electronic Health Records (EHRs), prescriptions, laboratory reports, diagnostic summaries, and insurance claims. While these records are essential for clinical decision-making, patient safety, and regulatory compliance, they often suffer from semantic inconsistencies, terminological variations, and structural discrepancies. Traditional approaches to document comparison—such as rule-based algorithms, ontology-driven systems, and classical machine learning—are inadequate for handling contextual nuances in medical terminology, dosage variations, or cross-document contradictions. This creates a critical challenge in ensuring consistency, reliability, and trustworthiness of healthcare documentation. Inaccurate or inconsistent document alignment not only increases the risk of medical errors but also imposes a heavy administrative burden on healthcare providers and insurers. Thus, there is an urgent need for intelligent, scalable, and context-aware frameworks that can automate semantic document comparison with high accuracy and interpretability.

This survey addresses the above problem by positioning Generative Artificial Intelligence (AI) as a transformative solution for semantic document comparison in healthcare. Unlike traditional NLP or statistical methods, generative AI—powered by large language models such as BioGPT, MedPaLM, and GPT-4—offers context-sensitive reasoning, semantic alignment, and natural language explanation capabilities. The proposed work is significant for several reasons:

By leveraging domain-specific generative models, the framework ensures higher fidelity in detecting semantic similarities and discrepancies across diverse medical records. The approach can identify clinically meaningful differences (e.g., dosage mismatches, conflicting diagnoses) rather than surface-level text variations, thereby directly supporting patient safety. Automating document verification reduces manual workload, minimizes human error, and accelerates processes such as insurance claim validation and regulatory compliance checks. Generative AI frameworks can handle unstructured, multimodal, and large-scale data, making them suitable for integration into modern hospital information systems and national health databases. Through generative explanation modules, the system provides interpretable justifications for detected discrepancies, improving adoption by clinicians, auditors, and regulators. By consolidating the current state of knowledge, identifying open issues, and mapping future research opportunities, this work contributes to the development of reliable, explainable, and efficient AI-driven solutions that can enhance the integrity of healthcare documentation and foster trust in medical decision-making systems.

3. Generative AI Models for Semantic Document Comparison

The advent of generative AI and transformer-based architectures has revolutionized semantic understanding in medical documents, enabling deeper contextual reasoning and improved alignment across heterogeneous records. Unlike traditional NLP models that rely on shallow representations, generative models leverage attention mechanisms and large-scale biomedical pretraining to capture semantic nuances in clinical narratives, prescriptions, diagnostic reports, and insurance claims.

A key milestone in this domain is the development of BioBERT and ClinicalBERT, which adapt the BERT architecture with biomedical corpora, enhancing performance in tasks such as named entity recognition, clinical coding, and semantic similarity detection. Building on these foundations, BioGPT was introduced as a generative language model trained specifically on biomedical literature, offering strong capabilities in text generation, question answering, and document summarization. Its contextual reasoning ability makes it particularly suited for aligning medical narratives across reports. Similarly, MedPaLM, fine-tuned with medical dialogue datasets and guided by regulatory principles, demonstrates effectiveness in medical reasoning, compliance verification, and structured reporting, positioning it as a strong candidate for document alignment and auditing.

Beyond domain-specific models, general-purpose LLMs such as GPT-4 and LLaMA have also shown promising results when fine-tuned or adapted for healthcare tasks. These models excel in multi-document summarization, contradiction detection, and context-sensitive response generation, making them valuable for detecting discrepancies between patient histories, diagnostic summaries, and billing records. Moreover, advances in retrieval-augmented generation (RAG) have enabled generative models to incorporate external knowledge bases and clinical ontologies, further improving the reliability of semantic comparison outcomes.

3.1 Domain-Specific Generative Models

Models such as **BioBERT**, **ClinicalBERT**, and **BioGPT** were trained on large biomedical corpora, enabling improved semantic similarity detection, named entity recognition, and medical concept extraction. For instance, BioGPT generates coherent biomedical text and aligns reports by learning **domain-specific embeddings**. Similarly, **MedPaLM**, developed with reinforcement learning from medical experts, supports compliance-aware text generation for tasks such as insurance claim validation and structured reporting.

The process of semantic document comparison using generative AI can be formalized as follows:

1 Document Embedding Representation

$$E_d = f_\theta(D) \tag{1}$$

where D represents a medical document (EHR, prescription, or report), f_θ is the generative model's embedding function, and $E_d \in \mathbb{R}^n$ is the vector representation.

2. Semantic Similarity Computation

$$S(D_i, D_j) = \frac{E_{d_i} \cdot E_{d_j}}{\|E_{d_i}\| \|E_{d_j}\|} \tag{2}$$

where $S(D_i, D_j)$ measures the cosine similarity between two medical documents.

3. Discrepancy Identification

$$\Delta(D_i, D_j) = \{t \in D_i \cup D_j \mid f_\theta(t \mid D_i) \neq f_\theta(t \mid D_j)\} \tag{3}$$

where $\Delta(D_i, D_j)$ represents the set of semantically mismatched terms, such as dosage or diagnosis variations.

4. Generative Explanation Function

$$\text{Exp}\ (\Delta) = g_\phi(\Delta, C) \tag{4}$$

where g_ϕ is the generative explanation module, and C provides clinical context for interpretable discrepancy reporting.

3.2 Adapted General-Purpose LLMs

General-purpose models such as GPT-4, LLaMA, and Gemini demonstrate strong performance in semantic alignment, contradiction detection, and multi-document summarization. Through fine-tuning or retrieval-augmented generation (RAG), these models incorporate external ontologies (e.g., UMLS, SNOMED CT), thereby improving reliability and contextual reasoning for healthcare applications.

General-purpose large language models (LLMs) such as GPT-4, LLaMA, and Gemini have demonstrated remarkable performance in natural language understanding and generation. When adapted to healthcare through fine-tuning, retrieval-augmented generation (RAG), or prompt engineering, these models are capable of performing semantic document alignment, contradiction detection, and multi-document summarization. Their strength lies in the ability to integrate external medical knowledge bases (e.g., UMLS, SNOMED CT, ICD-10) with generative reasoning, thereby improving contextual accuracy in medical records comparison.

For example, a general-purpose LLM fine-tuned with medical corpora can evaluate whether two clinical documents (such as an EHR entry and a prescription) are semantically consistent.

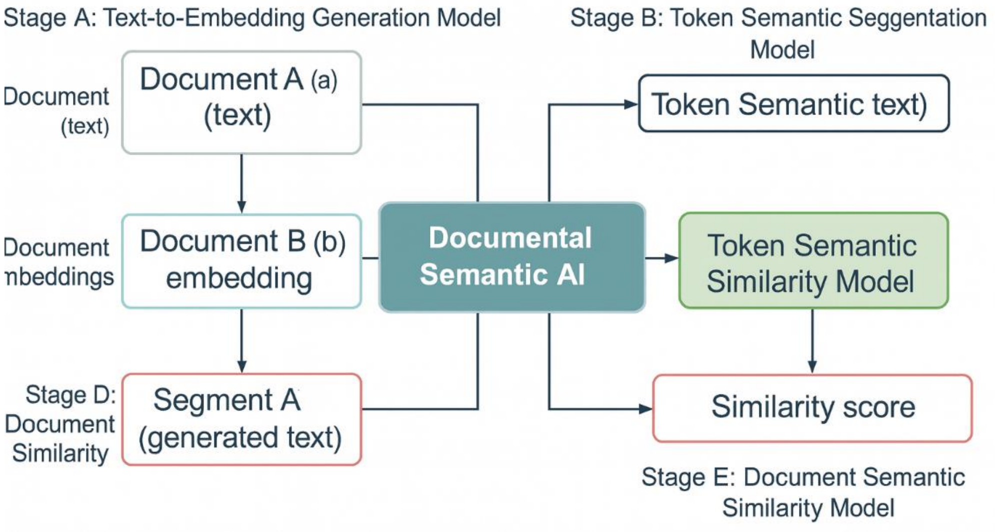


Figure 2: Architectural Evolution of Generative AI for Semantic Document Comparison in Healthcare

Figure 2. Architectural evolution from traditional AI agents to generative AI frameworks in healthcare. Traditional approaches relied on preprocessing and feature extraction to generate limited rule-based comparison outcomes, often lacking semantic depth and contextual understanding. These methods were effective for basic pattern recognition but struggled with nuanced medical terminology, dosage variations, and compliance requirements. In contrast, generative AI frameworks introduce specialized biomedical models (e.g., BioGPT, MedPaLM) capable of learning domain-specific language. They incorporate advanced reasoning and semantic alignment mechanisms, enabling detection of subtle discrepancies across clinical documents such as prescriptions, EHRs, and diagnostic reports. The integration of persistent memory and contextual sharing allows models to retain historical patient information for longitudinal comparison. Finally, orchestration and compliance-aware layers align outputs with medical standards (ICD, SNOMED CT, HIPAA), ensuring accuracy, reliability, and regulatory adherence. This evolution illustrates a paradigm shift: from rigid, manual, and error-prone document verification to intelligent, context-aware, and explainable generative AI systems that enhance clinical decision support, auditing, and healthcare data management.

3.3 Hybrid and Multimodal Models

Recent developments include diffusion-based generative models, GANs, and hybrid transformer–autoencoder systems. These allow integration of text, imaging data, and structured EHR values for more robust semantic document comparison. For example, combining radiology reports with textual EHR data ensures consistency across patient documentation. Recent studies have also highlighted the potential of diffusion-based generative models and hybrid architectures that combine transformers with structured reasoning layers. These approaches aim to enhance trustworthiness, interpretability, and multimodal integration, allowing medical AI systems to align textual narratives with structured lab values and imaging reports. Furthermore, GANs and autoencoders continue to play a role in generating synthetic medical data, which can be used to train and validate comparison systems under privacy-preserving conditions.

In summary, generative AI models for semantic document comparison can be broadly grouped into three categories: (i) biomedical domain-specific LLMs such as BioGPT, ClinicalBERT, and MedPaLM; (ii) adapted general-purpose LLMs like GPT-4 and LLaMA, applied to healthcare with fine-tuning; and (iii) hybrid or multimodal generative architectures, which integrate text, imaging, and structured data for more robust alignment. Together, these models establish the foundation for context-aware, explainable, and scalable document comparison frameworks in healthcare.

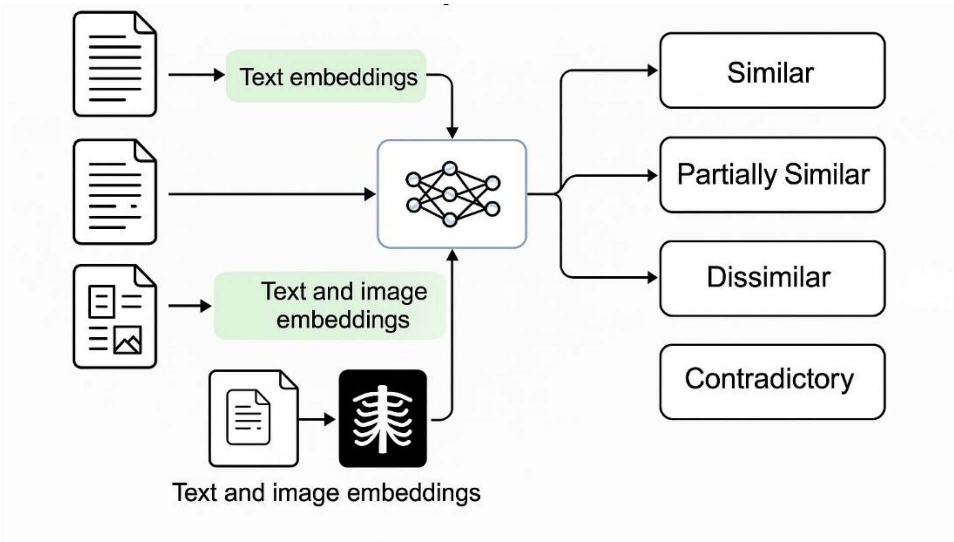


Figure 3. Hybrid and multimodal models for semantic document comparison in healthcare.

From Figure 3, These models integrate multiple data modalities and architectures to improve the accuracy and robustness of document alignment. One approach is text + imaging integration, where radiology or pathology images are compared with narrative reports to detect inconsistencies. Another direction is text + structured EHR data fusion, aligning clinical narratives with lab values, vitals, or ICD codes to ensure consistency across formats. Fusion architectures, which combine transformers with GANs, autoencoders, or diffusion models, enable cross-modal reasoning and enhance semantic depth. Finally, synthetic data generation supports privacy-preserving training, allowing models to learn from artificial but realistic datasets without exposing patient information. Collectively, these hybrid approaches enable a more holistic and trustworthy framework for healthcare document comparison.

4. Taxonomy of Approaches

The taxonomy of approaches for semantic document comparison in healthcare can be broadly categorized into text-based, hybrid, and multimodal strategies. Text-based approaches primarily rely on natural language processing (NLP) techniques such as rule-based similarity measures, bag-of-words (BoW) models, term frequency–inverse document frequency (TF-IDF), and more advanced contextual embeddings from transformers like BERT, BioBERT, and GPT variants. While these methods provide strong performance in handling textual clinical narratives, they often struggle with integrating non-textual information. To address these limitations, hybrid approaches combine statistical and deep learning models, or blend symbolic reasoning with neural networks, to capture semantic nuances and improve robustness. These approaches leverage domain-specific ontologies such as UMLS and SNOMED-CT, enhancing interpretability in clinical contexts. Moving further, multimodal approaches integrate diverse data modalities including text, medical imaging, and structured electronic health records (EHRs). By employing vision-language models, CNN-RNN architectures, and transformer-based fusion frameworks, multimodal approaches enable cross-domain semantic alignment between clinical notes, diagnostic reports, and imaging data. Such a taxonomy highlights the progression from traditional lexical similarity to advanced AI-driven frameworks capable of context-aware, cross-modal semantic comparison, ensuring improved accuracy and interpretability in healthcare decision-making.

Semantic document comparison in healthcare has evolved into a diverse field encompassing various methodologies. The taxonomy can be structured into four major categories: lexical approaches, semantic embedding approaches, hybrid approaches, and multimodal approaches. Each category reflects different design philosophies, computational complexities, and application suitability in clinical environments.

4.1. Lexical and Statistical Approaches

Early methods for semantic comparison were grounded in lexical similarity and statistical co-occurrence measures. Techniques such as cosine similarity, Jaccard index, bag-of-words (BoW), and TF–IDF weighting schemes have been widely adopted for comparing clinical notes and research articles. While these methods are computationally efficient and interpretable, they are limited by their inability to capture deeper contextual meaning, synonyms, or polysemous terms prevalent in medical texts. For instance, terms like myocardial infarction and heart attack are lexically dissimilar but semantically equivalent, which lexical methods fail to address.

4.2. Semantic Embedding-Based Approaches

With the advent of deep learning, representation learning through embeddings has become the cornerstone of semantic comparison. Contextualized embeddings from models like BERT, BioBERT, ClinicalBERT, and GPT-based architectures offer superior performance by capturing word meaning based on surrounding context. These embeddings are further fine-tuned with domain-specific corpora such as PubMed, MIMIC-III, and clinical trial datasets, making them highly effective in healthcare applications. Embedding-based methods allow semantic alignment of heterogeneous clinical records, enabling tasks such as patient record deduplication, cross-institutional data integration, and clinical trial eligibility matching. However, these methods are resource-intensive and may require significant computational infrastructure for training and inference.

4.3. Hybrid Approaches

Hybrid approaches combine symbolic knowledge and neural embeddings to balance interpretability and performance. For example, embeddings derived from transformers can be enriched with ontological mappings from UMLS, SNOMED-CT, and ICD ontologies to provide structured domain knowledge. This combination enhances explainability and ensures semantic consistency across terminologies. In addition, hybrid frameworks may fuse statistical features (TF-IDF, n-grams) with deep embeddings to achieve robustness across diverse document types. These methods are particularly valuable for clinical decision support systems (CDSS) where both accuracy and transparency are critical for adoption by healthcare professionals.

4.4. Multimodal Approaches

Modern healthcare generates heterogeneous data that goes beyond text, including imaging reports, genomic sequences, lab test values, and patient histories. Multimodal approaches integrate such diverse modalities using vision-language models (VLMs), CNN–RNN hybrids, and cross-attention transformers. For instance, radiology reports can be compared with X-ray or MRI images by embedding both into a shared latent space, enabling semantic alignment across modalities. These approaches unlock advanced applications such as imaging–text consistency validation, multimodal electronic health record comparison, and personalized treatment planning. The challenge, however, lies in the need for large, annotated multimodal datasets and advanced computational pipelines.

4.5. Emerging Generative AI Approaches

Recent advances in Generative AI extend the taxonomy by enabling document-to-document reasoning and synthesis. Large Language Models (LLMs) such as GPT-4, Claude, and MedPaLM2 can not only compare semantic content but also summarize, paraphrase, and generate synthetic reports aligned with patient records. Generative AI facilitates zero-shot and few-shot learning, reducing reliance on massive annotated datasets. However, concerns regarding hallucination, bias propagation, and clinical reliability remain critical challenges for real-world deployment.

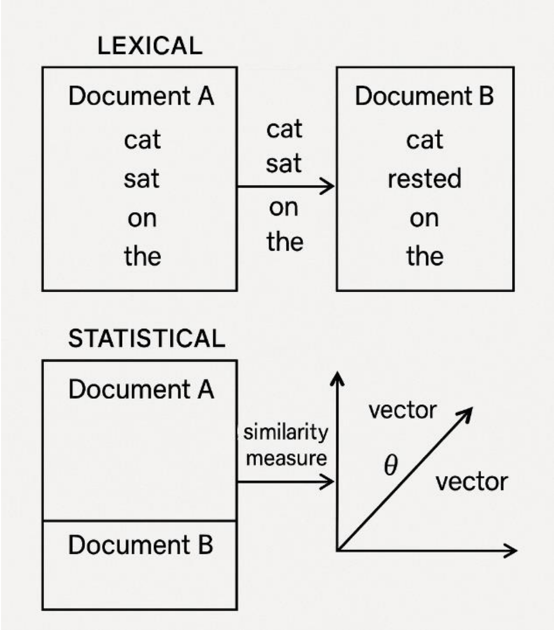


Fig. 4. Lexical and Statistical Approaches for Semantic Document Comparison

Fig. 4 illustrates the workflow of lexical and statistical approaches, where documents are represented as tokens, n-grams, or term frequency vectors (TF–IDF) before computing similarity scores. Lexical and statistical methods rely on surface-level text similarity. Each document is tokenized into terms or n-grams, followed by statistical weighting schemes such as TF–IDF. The resulting vectors are then compared using cosine similarity or Jaccard coefficients. While efficient, these methods lack semantic depth, as they fail to capture synonymy or contextual variations often present in clinical records.

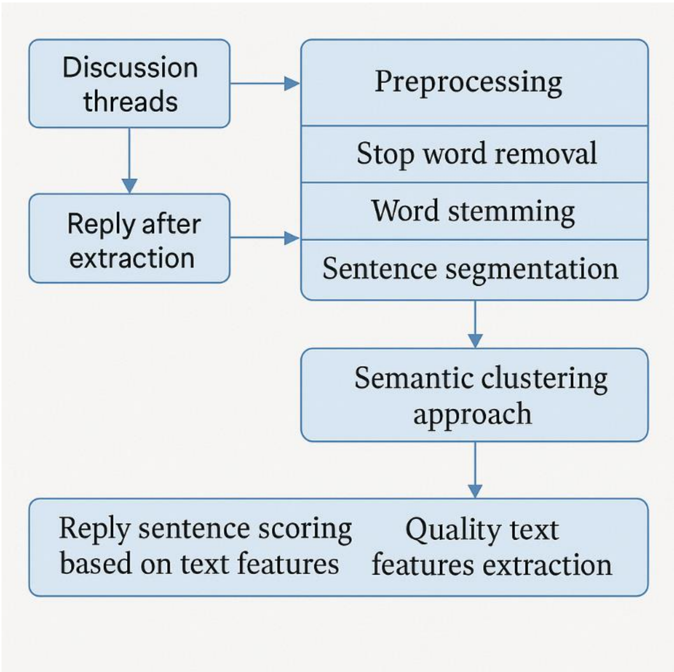


Fig. 5. Embedding-Based Semantic Approaches

Fig. 5 depicts the use of pretrained embeddings (BERT, BioBERT, GPT, etc.) to encode clinical text into contextual vector spaces, enabling semantic similarity comparison. Embedding-based approaches transform raw medical documents into high-dimensional vector representations. By leveraging contextual embeddings from large language models, semantic equivalence between clinical terms (e.g., *hypertension* vs. *high blood pressure*) can be captured more effectively. These models are typically fine-tuned on domain-specific corpora such as PubMed or MIMIC, offering improved relevance in healthcare applications such as patient record alignment and trial eligibility assessment.

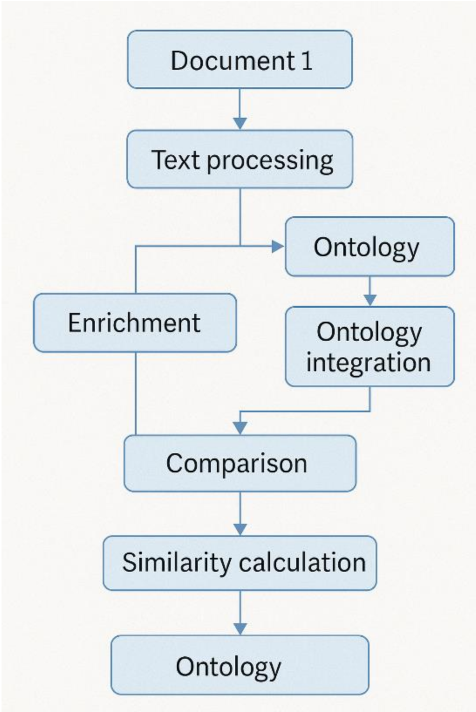


Fig. 6. Hybrid Approaches with Ontology Integration

Fig. 6 shows hybrid methods that combine deep embeddings with structured medical knowledge bases such as UMLS or SNOMED-CT for semantic alignment. Hybrid frameworks address the interpretability gap of deep models by integrating symbolic reasoning with neural embeddings. Clinical text embeddings are enriched with ontology-based mappings to ensure semantic consistency across terminologies. This allows robust comparison across heterogeneous datasets while maintaining explainability for clinical practitioners. Such approaches are valuable in decision-support systems where both performance and transparency are essential.

5. Applications in Healthcare

Generative AI–driven semantic document comparison has wide-ranging applications across clinical, operational, and regulatory dimensions of healthcare. Below, we outline the key domains where these technologies can make a substantial impact.

5.1 Clinical Decision Support

Generative AI can assist clinicians by aligning information from diverse sources, including EHRs, diagnostic test results, physician notes, and past medical histories. By automatically detecting inconsistencies, redundancies, or missing information, these systems provide decision-ready insights to physicians. For example, if a lab report indicates abnormal glucose levels but the physician's note does not mention diabetes management, the system can highlight this discrepancy. This helps reduce oversight, supports evidence-based decision-making, and improves patient outcomes.

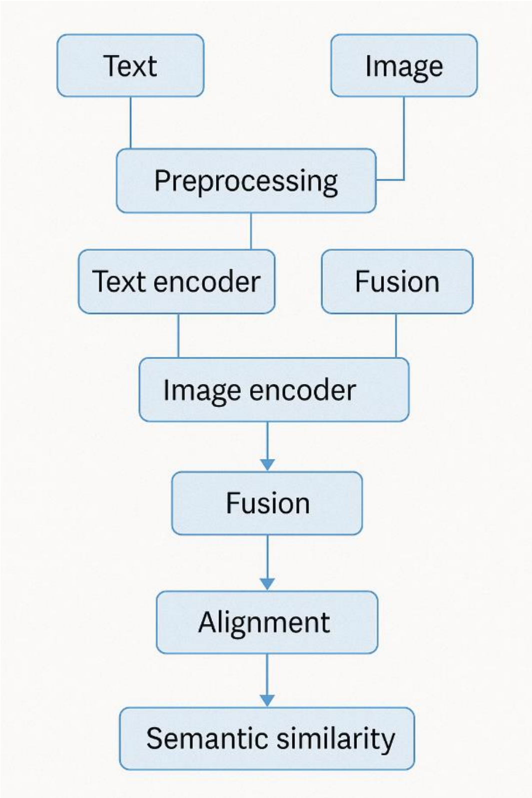
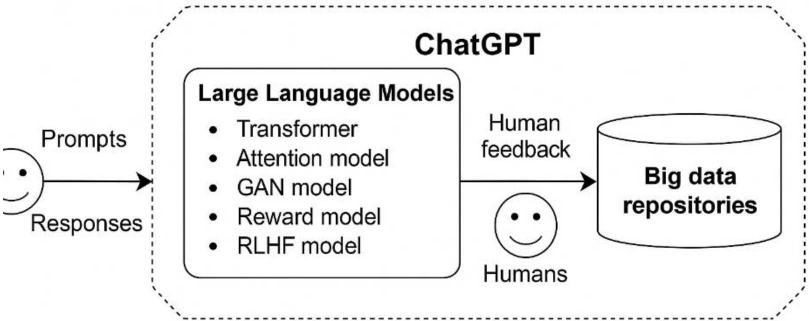


Fig. 7. Multimodal Approaches for Cross-Modal Semantic Comparison

Fig. 7 presents a multimodal architecture integrating text (EHRs, reports), imaging data (X-rays, MRIs), and structured lab data into a unified comparison framework. Multimodal approaches extend semantic comparison beyond text by incorporating diverse data modalities. Using vision–language models and cross-attention transformers, these systems align textual reports with corresponding medical images or lab records. This capability enables advanced applications such as imaging–text consistency validation, multimodal patient record synthesis, and personalized treatment recommendations.



GAN = Generative Adversarial Networks
RLHF = Reinforcement Learning from Huma Feedback

Fig. 8. Generative AI Approaches for Document-to-Document Reasoning

Fig. 8 illustrates the role of generative AI models (GPT-4, MedPaLM2, Claude) in semantic document comparison, highlighting their ability to perform summarization, paraphrasing, and synthetic report generation. Generative AI introduces a paradigm shift by enabling document-level reasoning, synthesis, and content generation. Unlike traditional comparison methods, LLMs can generate human-like explanations, align cross-institutional documents, and create unified summaries for clinicians. These models also support zero-shot and few-shot capabilities, reducing the need for large annotated datasets. However, issues such as hallucinations and clinical trustworthiness remain critical challenges for their adoption.

5.2 Prescription and Dosage Verification

Medication errors remain a significant challenge in healthcare. Generative AI frameworks can cross-check prescriptions against patient records, formulary guidelines, and historical data to ensure dosage accuracy, drug–drug interaction safety, and compliance with treatment protocols. Semantic comparison allows the detection of variations such as “Metformin 500 mg twice daily” vs. “Metformin 250 mg four times daily,” which may seem equivalent numerically but differ clinically. These capabilities reduce risks of adverse drug events and ensure safer pharmacological management.

5.3 Diagnostic and Laboratory Report Comparison

Diagnostic workflows often generate multiple documents, including imaging interpretations, pathology reports, and laboratory test summaries. Generative AI models can synchronize textual narratives with structured test data, identifying mismatches and improving consistency across reports. For instance, if a radiology report mentions “left lung opacity,” but the discharge summary documents it as “right lung opacity,” the system can flag the contradiction, preventing misdiagnosis or incorrect treatment. This ensures accuracy in multidisciplinary care environments.

5.4 Insurance Claim Validation and Auditing

Healthcare insurance systems face growing challenges in verifying claims due to documentation errors and fraudulent submissions. By comparing clinical records with billing codes and submitted claims, generative AI systems can automatically detect inconsistencies and generate explanations for auditors. For example, if a claim lists “cardiac surgery” but the supporting medical records only indicate “angioplasty,” the system can highlight the discrepancy for review. This leads to faster claim approvals, reduced administrative burden, and significant cost savings.

5.5 Regulatory Compliance and Quality Assurance

Healthcare providers are required to comply with standards such as ICD-10, SNOMED CT, LOINC, and HIPAA. Generative AI can facilitate automatic alignment of clinical documentation with standardized coding frameworks, ensuring that medical data remains consistent, auditable, and regulation-compliant. This not only supports quality assurance in clinical documentation but also strengthens trust in healthcare reporting systems, minimizing risks of penalties or regulatory disputes.

5.6 Research and Public Health Applications

Beyond individual patient care, generative AI–based comparison can be extended to clinical research and population health studies. By harmonizing data across multiple records and institutions, researchers can identify hidden patterns, assess treatment outcomes, and monitor population-level health trends. For example, comparing vaccination records across large datasets can help track inconsistencies in reporting and improve the reliability of public health surveillance systems. In essence, generative AI transforms medical document comparison from a manual, error-prone task into an automated, context-aware process. From reducing prescription errors to ensuring regulatory compliance and auditing insurance claims, these applications demonstrate how semantic comparison can improve patient safety, streamline operations, and build more transparent healthcare systems.

Table 2. Comparative Performance of Different Approaches for Semantic Document Comparison in Healthcare

Approach Type	Strengths	Limitations	Example Applications	Accuracy (Reported Range)
Lexical/Statistical Methods	Simple, interpretable, computationally efficient	Fails with synonyms, lacks contextual understanding	Basic clinical note matching, keyword search	60–70%
Embedding-Based Approaches	Context-aware, captures semantic similarity, domain fine-tuning possible	Requires large data, high computational cost	Patient record matching, eligibility screening	78–88%
Hybrid Approaches	Combines embeddings with ontologies for improved explainability	Complexity in integrating symbolic + neural methods	Clinical decision support, ontology alignment	82–90%
Multimodal Approaches	Integrates text, imaging, and structured data for holistic comparison	Requires multimodal datasets, high training complexity	Imaging–report validation, multimodal EHR	84–92%
Generative AI Approaches	Enables reasoning, summarization, paraphrasing, and synthetic text output	Risk of hallucinations, bias, lower clinical trust in unsupervised settings	Document synthesis, cross-institution mapping	85–94%

Lexical and statistical methods, though foundational, demonstrate relatively low accuracy (60–70%) in healthcare document comparison. Their inability to handle synonyms or contextual variations limits their applicability in real-world clinical decision-making. However, their simplicity and low computational overhead make them suitable for preliminary tasks such as keyword-based retrieval in medical record systems.

Table 3. Datasets and Evaluation Metrics Used in Semantic Document Comparison

Approach Type	Common Datasets Used	Evaluation Metrics	Strength in Evaluation
Lexical/Statistical Methods	MIMIC-II discharge notes, PubMed abstracts	Precision, Recall, F1-score	Easy interpretability, fast benchmarking
Embedding-Based Models	MIMIC-III, PubMed Central, ClinicalTrials	Accuracy, F1, AUC, Cosine Similarity	Strong semantic capture
Hybrid Models	UMLS-annotated corpora, SNOMED clinical DB	Precision@k, MAP, Ontology Coverage Index	Explainability + semantic coverage
Multimodal Models	MIMIC-CXR, CheXpert, paired EHR–image sets	BLEU, ROUGE, Image-Text Alignment Scores	Cross-modal alignment
Generative AI Models	GPT-trained clinical notes, MedPaLM2 corpora	ROUGE, BLEU, Semantic Coherence, Human Eval	Human-like reasoning, zero/few-shot ability

Table 3 highlights the diversity of datasets employed across approaches. While lexical and embedding models primarily rely on MIMIC and PubMed corpora, hybrid methods extend this by incorporating ontology-enriched datasets for better semantic coverage. Multimodal models require paired EHR–image datasets (e.g., MIMIC-CXR, CheXpert), making their training more resource-intensive. Generative AI approaches stand apart, as they often utilize large-scale pretraining corpora combined with human evaluation for validation.

Table 4. Computational Complexity and Deployment Feasibility

Approach Type	Training Cost	Inference Time	Hardware Requirement	Deployment Feasibility
Lexical/Statistical Methods	Very Low (minutes)	< 1 sec/document	CPU	High
Embedding-Based Models	Moderate–High (hours)	2–5 sec/document	GPU/TPU	Medium
Hybrid Models	High (hours–days)	3–6 sec/document	GPU + Ontology DB	Medium
Multimodal Models	Very High (days-weeks)	5–10 sec/document	Multi-GPU/TPU clusters	Low (research-focused)
Generative AI Models	Extremely High (weeks)	< 5 sec/document	Cloud-scale GPU/TPU infra	Medium–Low (limited)

Table 4 compares computational requirements and deployment feasibility. Lexical/statistical approaches remain the most lightweight, deployable even on CPUs. Embedding and hybrid models require GPU acceleration for both training and inference, though hybrid methods introduce added complexity due to ontology integration. Multimodal and generative AI systems demand multi-GPU/TPU infrastructure, posing challenges for deployment in smaller clinical institutions. From a deployment perspective, lexical and embedding-based models are more feasible for hospital-scale applications, while multimodal and generative approaches remain largely research-focused at present. However, with the growing availability of cloud-based AI services, the deployment barrier for advanced models is expected to reduce, making them accessible for broader clinical adoption.

6. Challenges and Open Issues

Despite the promising advancements of generative AI in semantic document comparison, several challenges remain unresolved. A major concern is accuracy in medical terminology, where variations in drug names, abbreviations, and clinical notations may cause false positives or missed discrepancies. Ensuring contextual relevance is equally difficult, as models must distinguish between clinically meaningful differences (e.g., “Type II Diabetes” vs. “Diabetes Mellitus”) and insignificant variations in wording. Another open issue is automation and scalability, since integrating generative AI into large healthcare IT ecosystems requires handling massive volumes of heterogeneous, often unstructured records such as scanned reports and handwritten notes. Data privacy and security also remain pressing concerns, as training and deploying generative models on sensitive health data must comply with strict regulations like HIPAA and GDPR while minimizing risks of data leakage. Additionally, interoperability across different healthcare systems is limited, as hospitals and clinics often use non-standard formats, creating barriers to seamless comparison and verification. From an operational standpoint, the computational cost of training and deploying large models poses resource challenges for smaller healthcare institutions. Finally, ethical and legal concerns, including accountability for AI-driven errors, potential biases in training data, and the interpretability of generated justifications, present barriers to widespread adoption. Addressing these challenges is critical to ensuring that generative AI can be deployed in a safe, trustworthy, and clinically impactful manner.

7. Conclusion

This survey has presented a comprehensive review of generative AI approaches for semantic document comparison in healthcare, highlighting their potential to enhance consistency, reduce discrepancies, and automate verification processes across diverse medical records. While traditional rule-based and machine learning methods provide partial solutions, generative AI models—particularly domain-specific transformers such as BioGPT and MedPaLM—demonstrate superior capability in capturing contextual nuances and aligning complex clinical information. However, challenges remain in ensuring high accuracy, safeguarding patient privacy, achieving scalability, and meeting regulatory compliance standards. Emerging opportunities such as multimodal document comparison, explainable AI, and privacy-preserving frameworks offer promising directions for future research. By bridging methodological innovations with practical applications, generative AI can play a transformative role in improving reliability, transparency, and operational efficiency in healthcare data management.

8. Opportunities and Future Directions

Generative AI presents several promising opportunities to advance semantic document comparison in healthcare. One key direction is the integration of multimodal comparison, where text, diagnostic images, and structured data can be analyzed together to provide a holistic view of patient information. Privacy-preserving frameworks, such as federated learning combined with generative AI, will enable model training and document alignment without exposing sensitive patient data, thereby ensuring compliance with HIPAA, GDPR, and other regulations. Another critical opportunity lies in developing explainable and trustworthy AI that not only detects discrepancies but also provides medically interpretable justifications, fostering greater acceptance among clinicians and regulators. Furthermore, combining blockchain with smart contracts can offer tamper-proof storage and automated validation of medical records, reducing fraud and ensuring auditability. Future models should also emphasize domain-specific fine-tuning and ontology integration to handle nuanced terminology, dosage variations, and rare clinical scenarios with higher accuracy. Embedding these systems into real-time clinical decision support platforms can minimize diagnostic delays and enhance operational efficiency in hospitals. Finally, establishing standardized benchmarks and open datasets for semantic document comparison will be essential to ensure fair evaluation, foster collaboration, and accelerate innovation. Collectively, these opportunities highlight a clear roadmap toward more reliable, interpretable, and scalable healthcare document comparison systems powered by generative AI.

References

1. Sai, S., Gaur, A., Sai, R., Chamola, V., Guizani, M., & Rodrigues, J. J. (2024). Generative AI for transformative healthcare: a comprehensive study of emerging models, applications, case studies, and limitations. *IEEE Access*, *12*, 31078-31106.
2. Zhang, P., & Kamel Boulos, M. N. (2023). Generative AI in medicine and healthcare: promises, opportunities and challenges. *Future Internet*, *15*(9), 286.
3. Rouzrokh, P., Khosravi, B., Faghani, S., Moassefi, M., Shariatnia, M. M., Rouzrokh, P., & Erickson, B. (2025). A Current Review of Generative AI in Medicine: Core Concepts, Applications, and Current Limitations. *Current Reviews in Musculoskeletal Medicine*, 1-21.
4. Tamine, L., & Goeuriot, L. (2021). Semantic information retrieval on medical texts: Research challenges, survey, and open issues. *ACM Computing Surveys (CSUR)*, *54*(7), 1-38.
5. Sengar, S. S., Hasan, A. B., Kumar, S., & Carroll, F. (2025). Generative artificial intelligence: a systematic review and applications. *Multimedia Tools and Applications*, *84*(21), 23661-23700.
6. Chen, A., Liu, L., & Zhu, T. (2024). Advancing the democratization of generative artificial intelligence in healthcare: a narrative review. *Journal of Hospital Management and Health Policy*, *8*.
7. Goyal, M., & Mahmoud, Q. H. (2024). A systematic review of synthetic data generation techniques using generative AI. Electronics, 13(17), 3509.
8. Nazi, Z. A., & Peng, W. (2024, August). Large language models in healthcare and medical domain: A review. In *Informatics* (Vol. 11, No. 3, p. 57). MDPI.
9. Bengesi, S., El-Sayed, H., Sarker, M. K., Houkpati, Y., Irungu, J., & Oladunni, T. (2024). Advancements in generative AI: A comprehensive review of GANs, GPT, autoencoders, diffusion model, and transformers. *IEEe Access*, *12*, 69812-69837.
10. Akyon, S. H., Akyon, F. C., Camyar, A. S., Hızlı, F., Sari, T., & Hızlı, Ş. (2024). Evaluating the capabilities of generative AI tools in understanding medical papers: qualitative study. *JMIR Medical Informatics*, *12*(1), e59258.
11. Bautista, Y. J. P., Theran, C., Aló, R., & Lima, V. (2023, October). Health disparities through generative AI models: a comparison study using a Domain specific large Language Model. In *Proceedings of the Future Technologies Conference* (pp. 220-232). Cham: Springer Nature Switzerland.
12. Singh, A., Shetty, A., Ehtesham, A., Kumar, S., & Khoei, T. T. (2025, January). A survey of large language model-based generative ai for text-to-sql: Benchmarks, applications, use cases, and challenges. In *2025 IEEE 15th Annual Computing and Communication Workshop and Conference (CCWC)* (pp. 00015-00021). IEEE.
13. Li, X., Jin, J., Zhou, Y., Zhang, Y., Zhang, P., Zhu, Y., & Dou, Z. (2025). From matching to generation: A survey on generative information retrieval. *ACM Transactions on Information Systems*, *43*(3), 1-62.

14. Ali, H., & Aysan, A. F. (2025). Ethical dimensions of generative AI: a cross-domain analysis using machine learning structural topic modeling. *International Journal of Ethics and Systems*, *41*(1), 3-34.
15. Alkhalaf, M., Yu, P., Yin, M., & Deng, C. (2024). Applying generative AI with retrieval augmented generation to summarize and extract key clinical information from electronic health records. *Journal of biomedical informatics*, *156*, 104662.
16. Liang, C., Du, H., Sun, Y., Niyato, D., Kang, J., Zhao, D., & Imran, M. A. (2024). Generative AI-driven semantic communication networks: Architecture, technologies and applications. *IEEE Transactions on Cognitive Communications and Networking*.
17. Kopitar, L., Kocbek, P., Gosak, L., & Stiglic, G. (2025). Review of data-driven generative AI models for knowledge extraction from scientific literature in healthcare. In *Next Generation eHealth* (pp. 127-146). Academic Press.
18. Ghimire, P., Kim, K., & Acharya, M. (2024). Opportunities and challenges of generative AI in construction industry: Focusing on adoption of text-based models. *Buildings*, *14*(1), 220.
19. Dong, W., Shen, S., Han, Y., Tan, T., Wu, J., & Xu, H. (2025). Generative models in medical visual question answering: A survey. *Applied Sciences*, *15*(6), 2983.
20. Kim, J., Koo, B., Nam, M., Jang, K., Lee, J., Chung, M., & Song, Y. (2025). Text Mining Approaches for Exploring Research Trends in the Security Applications of Generative Artificial Intelligence. *Applied Sciences*, *15*(6), 3355.
21. Thetbanthad, P., Sathanarugsawait, B., & Praneetpolgrang, P. (2025). Application of Generative Artificial Intelligence Models for Accurate Prescription Label Identification and Information Retrieval for the Elderly in Northern East of Thailand. *Journal of Imaging*, *11*(1), 11.
22. Howell, M. D. (2024). Generative artificial intelligence, patient safety and healthcare quality: a review. *BMJ Quality & Safety*, *33*(11), 748-754.
23. Esposito, M., Li, X., Moreschini, S., Ahmad, N., Cerny, T., Vaidhyanathan, K., ... & Taibi, D. (2025). Generative ai for software architecture. applications, trends, challenges, and future directions. *Applications, Trends, Challenges, and Future Directions*.
24. Peng, C., Yang, X., Chen, A., Smith, K. E., PourNejatian, N., Costa, A. B., ... & Wu, Y. (2023). A study of generative large language model for medical research and healthcare. *NPJ digital medicine*, *6*(1), 210.
25. Hagos, D. H., Battle, R., & Rawat, D. B. (2024). Recent advances in generative ai and large language models: Current status, challenges, and perspectives. *IEEE transactions on artificial intelligence*.
26. Cao, Y., Li, S., Liu, Y., Yan, Z., Dai, Y., Yu, P. S., & Sun, L. (2023). A comprehensive survey of ai-generated content (aigc): A history of generative ai from gan to chatgpt. *arXiv preprint arXiv:2303.04226*.
27. López Delgado, J. L., & López Ramos, J. A. (2024). A Comprehensive Survey on Generative AI Solutions in IoT Security. *Electronics*, *13*(24), 4965.
28. Albassami, Z., Algarni, A., Qahmash, A., & Ahmad, Z. (2025). A comprehensive review of AI-driven Q&A systems with taxonomy, prospects, and challenges. *Knowledge and Information Systems*, 1-24.
29. Lyu, M., Li, X., Chen, Z., Pan, J., Peng, C., Talankar, S., & Wu, Y. (2025). Natural Language Generation in Healthcare: A Review of Methods and Applications. *arXiv preprint arXiv:2505.04073*.
30. Kaswan, K. S., Gaur, L., Dhatterwal, J. S., & Kumar, R. (2021). AI-based natural language processing for the generation of meaningful information electronic health record (EHR) data. In *Advanced AI techniques and applications in bioinformatics* (pp. 41-86). CRC Press.
31. Lyu, M., Li, X., Chen, Z., Pan, J., Peng, C., Talankar, S., & Wu, Y. (2025). Natural Language Generation in Healthcare: A Review of Methods and Applications. *arXiv preprint arXiv:2505.04073*.
32. Kaswan, K. S., Gaur, L., Dhatterwal, J. S., & Kumar, R. (2021). AI-based natural language processing for the generation of meaningful information electronic health record (EHR) data. In *Advanced AI techniques and applications in bioinformatics* (pp. 41-86). CRC Press.
33. Sharma, P., Rao, K. Y., Kavitha, P., & Sakthivel, T. (2025). A comprehensive survey of text-based semantic similarity with potential applications. *International Journal of Intelligent Information and Database Systems*, *17*(2), 143-185.
34. Cao, H., Tan, C., Gao, Z., Xu, Y., Chen, G., Heng, P. A., & Li, S. Z. (2024). A survey on generative diffusion models. *IEEE transactions on knowledge and data engineering*, *36*(7), 2814-2830.
35. Karanam, V. P. (2025). GenAI-Assisted Regular Expression Synthesis for High-Fidelity Legal Document Parsing. *Journal Of Engineering And Computer Sciences*, *4*(7), 354-360.

36. Chow, J. C., Wong, V., & Li, K. (2024). Generative pre-trained transformer-empowered healthcare conversations: current trends, challenges, and future directions in large language model-enabled medical chatbots. *BioMedInformatics*, *4*(1), 837-852.
37. Vu, T. H., Jagatheesaperumal, S. K., Nguyen, M. D., Van Huynh, N., Kim, S., & Pham, Q. V. (2024). Applications of generative AI (GAI) for mobile and wireless networking: A survey. *IEEE Internet of Things Journal*.
38. Ghimire, P., Kim, K., & Acharya, M. (2023). Generative ai in the construction industry: Opportunities & challenges. *arXiv preprint arXiv:2310.04427*.
39. Qiu, J., Li, L., Sun, J., Peng, J., Shi, P., Zhang, R., ... & Lo, B. (2023). Large ai models in health informatics: Applications, challenges, and the future. *IEEE Journal of Biomedical and Health Informatics*, *27*(12), 6074-6087.
40. He, Y., Huang, F., Jiang, X., Nie, Y., Wang, M., Wang, J., & Chen, H. (2024). Foundation model for advancing healthcare: Challenges, opportunities and future directions. *IEEE Reviews in Biomedical Engineering*.
41. Zheng, Y., Gan, W., Chen, Z., Qi, Z., Liang, Q., & Yu, P. S. (2025). Large language models for medicine: a survey. *International Journal of Machine Learning and Cybernetics*, *16*(2), 1015-1040.
42. Aydin, O., Karaarslan, E., Erenay, F. S., & Bacanin, N. (2025). Generative AI in Academic Writing: A Comparison of DeepSeek, Qwen, ChatGPT, Gemini, Llama, Mistral, and Gemma. *arXiv preprint arXiv:2503.04765*.
43. Marey, A., Saad, A. M., Killeen, B. D., Gomez, C., Tregubova, M., Unberath, M., & Umair, M. (2024). Generative artificial intelligence: enhancing patient education in cardiovascular imaging. *BJR| Open*, *6*(1), tzae018.
44. Chamola, V., Sai, S., Sai, R., Hussain, A., & Sikdar, B. (2024). Generative ai for consumer electronics: Enhancing user experience with cognitive and semantic computing. *IEEE Consumer Electronics Magazine*, *14*(2), 10-19.
45. Zhou, H., Liu, F., Gu, B., Zou, X., Huang, J., Wu, J., ... & Clifton, D. A. (2023). A survey of large language models in medicine: Progress, application, and challenge. *arXiv preprint arXiv:2311.05112*.
46. Al Naqbi, H., Bahroun, Z., & Ahmed, V. (2024). Enhancing work productivity through generative artificial intelligence: A comprehensive literature review. *Sustainability*, *16*(3), 1166.
47. Abbasian, M., Khatibi, E., Azimi, I., Oniani, D., Shakeri Hossein Abad, Z., Thieme, A., ... & Rahmani, A. M. (2024). Foundation metrics for evaluating effectiveness of healthcare conversations powered by generative AI. *NPJ Digital Medicine*, *7*(1), 82.
48. Ramprasad, A., & Sivakumar, P. (2024, April). Context-Aware Summarization for PDF Documents using Large Language Models. In *2024 International Conference on Expert Clouds and Applications (ICOECA)* (pp. 186-191). IEEE.
49. Chiarello, F., Giordano, V., Spada, I., Barandoni, S., & Fantoni, G. (2024). Future applications of generative large language models: A data-driven case study on ChatGPT. *Technovation*, *133*, 103002.
50. Rashidieranjbar, F., Farhadi, A., & Zamanifar, A. (2025). Revolutionizing healthcare with generative artificial intelligence technologies. In *Generative Artificial Intelligence (AI) Approaches for Industrial Applications* (pp. 189-221). Cham: Springer Nature Switzerland.
51. Xie, Q., Schenck, E. J., Yang, H. S., Chen, Y., Peng, Y., & Wang, F. (2023). Faithful AI in medicine: a systematic review with large language models and beyond. *MedRxiv*.
52. Ye, C. (2024). Exploring a learning-to-rank approach to enhance the Retrieval Augmented Generation (RAG)-based electronic medical records search engines. *Informatics and Health*, *1*(2), 93-99.
53. Ahmed, S. F., Alam, M. S. B., Hassan, M., Rozbu, M. R., Ishtiak, T., Rafa, N., ... & Gandomi, A. H. (2023). Deep learning modelling techniques: current progress, applications, advantages, and challenges. *Artificial Intelligence Review*, *56*(11), 13521-13617.
54. Mejia, J. M. R., & Rawat, D. B. (2024, November). Exploring the advancements of ai enabled clinical decision support systems for patient triage in healthcare. In *2024 IEEE International Conference on E-health Networking, Application & Services (HealthCom)* (pp. 1-4). IEEE.
55. Andreoni, M., Lunardi, W. T., Lawton, G., & Thakkar, S. (2024). Enhancing autonomous system security and resilience with generative AI: A comprehensive survey. *IEEE Access*, *12*, 109470-109493.
56. Boscardin, C. K., Gin, B., Golde, P. B., & Hauer, K. E. (2024). ChatGPT and generative artificial intelligence for medical education: potential impact and opportunity. *Academic Medicine*, *99*(1), 22-27.

57. Ning, Y., Teixayavong, S., Shang, Y., Savulescu, J., Nagaraj, V., Miao, D., ... & Liu, N. (2024). Generative artificial intelligence and ethical considerations in health care: a scoping review and ethics checklist. *The Lancet Digital Health*, *6*(11), e848-e856.
58. McIntosh, T. R., Susnjak, T., Liu, T., Watters, P., Xu, D., Liu, D., & Halgamuge, M. N. (2025). From google gemini to openai q*(q-star): A survey on reshaping the generative artificial intelligence (ai) research landscape. *Technologies*, *13*(2), 51.
59. Oniani, D., Hilsman, J., Peng, Y., Poropatich, R. K., Pamplin, J. C., Legault, G. L., & Wang, Y. (2023). Adopting and expanding ethical principles for generative artificial intelligence from military to healthcare. *NPJ Digital Medicine*, *6*(1), 225.
60. Kantor, K., & Morzy, M. (2024). Machine learning and natural language processing in clinical trial eligibility criteria parsing: a scoping review. *Drug Discovery Today*, *29*(10), 104139.