Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 08:2025
DOI: 10.5281/zenodo.16932563

GENERATIVE AI INTEGRATION FOR REAL-TIME PATIENT DATA INTERPRETATION: A COMPREHENSIVE PERFORMANCE ANALYSIS AND CLINICAL IMPLEMENTATION FRAMEWORK

NCHEBE-JAH RAYMOND ILOANUSI

M.D, College of Sten Island, 2800 Victory BLVD, Staten Island, NY.
Email: nchebejah.iloanusi@csi.cuny.edu, ORCID: https://orcid.org/0009-0009-4224-1383

NZUBE-JAH UKAH

Bachelor of Pharmacy (B.Pharm), University of Western Ontario, 1151 Richmond Street, N6A3K7, London, Ontario. Email: nukah@uwo.ca, ORCID: https://orcid.org/0009-0007-7545-9405

AMARACHI CONFIDENCE NWEKE

B.Sc. Nursing, Elizade University, Ilara Mokin, 340112, Ondo, Nigeria.
Email: confidenceamarachinweke@gmail.com

Abstract

Real-time interpretation of complex patient data simplifies the clinical processes with Generative Artificial Intelligence (AI). This experiment compares the performance of three generative artificial intelligence models — ChatGPT, Julius, and Claude — with that of an expert biomedical model, BioBERT, on over 4,000 patient comments regarding antidiabetic drugs. The data were preprocessed and anonymized, and then the sentiment was analyzed, along with the identification of adverse effects, therapeutic outcomes, and thematic classification. Modes were ranked based on accuracy, depth of interpretation, clinical relevance, ability to process and produce actionable insights, and speed. Findings reveal that the two general-purpose models outperformed BioBERT when evaluating their performance in terms of narrative generation and contextual reasoning, whereas BioBERT has surpassed the general-purpose models when tested on recognition of medical terms, pharmacology accuracy, and adverse event detection. These results reveal the foundations of flexibility in lieu of clinical precision trade-offs and the potential for possessing hybrid AI through moves that combine the conveniences. A clinical implementation proposal is submitted, including details on how and where it should integrate with electronic health records (EHR), its compliance with regulations, and how it can be utilized in the training of healthcare professionals and in communicating with the general population. This is a practical exposition on the use of generative AI to enhance patient care and operational delivery in managing diabetes.

Keywords: Generative AI in Healthcare, Real-Time Clinical Data Analysis, Patient-Generated Health Data, Endocrinology Informatics, AI-Assisted Diagnosis, Electronic Health Record Integration, Antidiabetic Medication Analysis, Clinical Natural Language Processing.

1. INTRODUCTION

AI is enhancing the medical field by enabling it to learn, train, and analyze complex clinical information (Zhang & Kamel Boulos, 2023; Sai et al., 2024; Yu et al., 2023). Patient data in real-time also offers specific possibilities for analysis with the help of generative AI, which has the potential to process information into clear, relevant answers tailored to a particular case (Nova, 2023; Reddy, 2024). It is also capable of enhancing natural language processing (Wu et al., 2020; Gao et al., 2023) of both structured and unstructured health data, which can be based on large language models and customized

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 08:2025
DOI: 10.5281/zenodo.16932563

training. The new wave of patient-generated health data (PGHD) obtained through wearable devices, mobile applications, and patient communities promises to contain gold mines that have never been exploited, i.e., information regarding patients and how they feel during treatment (Abdolkhani et al., 2019; Tiase et al., 2020; Lordon et al., 2020).

The unstructured format, heterogeneity of quality, and implications of integration are the factors that explain the inability to integrate PGHD into clinical practice (Kawu et al., 2023; Winter & Davidson, 2022). It requires the presence of interoperability and governance standards because the absorption of PGHD within the electronic health record (EHR) should be done in a privacy-friendly manner (Nittas et al., 2019; Shaw et al., 2020).

Generative AI has the potential to convert heterogeneous data to clinical use in terms of clinical inference, sentiment analysis, and actionable summary (Sheikhalishahi et al., 2019; Soysal et al., 2018; Michelson et al., 2020) and has been shown to be aligned with the initiatives in the direction of personalization (Hawley et al., 2021; Mahajan et al., 2023). Notwithstanding, a number of technical and adoption barriers to full-scale adoption exist, e.g., the exploitation of federated learning (Camajori Tedeschini et al., 2022) and trust issues among clinicians (Cheng et al., 2022; Hsieh, 2023).

It benchmarks generative AI models on interpreting over 4,000 patient comments about antidiabetic medications against domain-specific biomedical AI models, measuring accuracy, clinical relevance, and speed. It even suggests an implementation framework in clinical settings regarding the idea of integrating EHRs, regulatory management, and public health usages (Demner-Fushman et al., 2009; Dinh-Le et al., 2019; Sharifshazileh et al., 2021).

2. DATASET DESCRIPTION AND PREPROCESSING

The data used in this paper comprise more than 4,000 comments posted by patients regarding various antidiabetic medications. Patient-generated health data (PGHD) has been cited as an important source of real-world evidence for developing better care provision, pharmacovigilance, and individualized treatment plans (Abdolkhani et al., 2019; Lordon et al., 2020; Nittas et al., 2019; Tiase et al., 2020). One of the issues affected by the consideration of secondary data is that there are online spaces where individuals can share their experiences with medication, which is public and raises concerns about the ethics of conducting such research.

Given the sensitivity of PGHD, anonymization measures were implemented to deprive the patients of the possibility of being recognized by their personally identifiable information based on the requirements of HIPAA and GDPR, which offer the best practices in terms of governance and socio-technical aspects of EHR-integrated datasets (Kawu et al., 2023; Winter & Davidson, 2022; Yu et al., 2023).

This also renders the data suitable for training an AI model in a clinical setting without infringing on patient privacy and violating relevant regulatory policies.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 08:2025
DOI: 10.5281/zenodo.16932563

The comments were prepared to accommodate generative AI and NLP-specific medical models, ideally. In the preprocessing process, the clinical NLP pipelines were followed (Wu et al., 2020; Sheikhalishahi et al., 2019; Soysal et al., 2018), which were as follows:

- **Noise Solution:** Providing removal of extraneous metadata, duplicates, as well as Web-based data remnants, and maintaining the integrity of the information data (Michelson et al., 2020).
- **Tokenization:** This is the process of breaking down the text into enabling sections that are analyzed into sentences and words, facilitating the analysis of the syntax and semantics of the text across models (Demner-Fushman et al., 2009).
- **Sentiment Labeling:** This approach would categorize every comment by labeling it as positive, negative, or neutral regarding therapeutic effects and side effects (Gao et al., 2023; Sharifshazileh et al., 2021).
- **Domain Specific Entity Recognition:** A means of recognizing regulatory agreement between the signal of drugs, dosages, and side effects with the regulatory medical ontologies (Yao et al., 2023).

The preprocessing was key to the possibility of computationally measuring the relative performance of both a general-purpose generative AI model (Julius, Claude) and a domain-specific medical AI model (BioBERT). Available literature has demonstrated that the quality of preprocessing can improve the interpretability and accuracy of AI, particularly in cases where PGHD can supplement existing EHR systems (Shaw et al., 2020; Dinh-Le et al., 2019; Mahajan et al., 2023).

Table 1: Dataset Description and Preprocessing

Characteristic	Value	Description
Total Comments	4,012	Number of individual patient-generated medication reviews.
Average Comment Length	45 words	Calculated after noise reduction and tokenization.
Sentiment Distribution	54% Positive, 28% Negative, 18% Neutral	Annotated manually and validated by AI models.
Medication Categories	12	Includes metformin, insulin analogs, DPP-4 inhibitors, SGLT2 inhibitors, etc.
Data Source Platforms	5	Public health forums, review sites, and patient advocacy group discussions.
Anonymization Compliance Frameworks	HIPAA, GDPR	Removed all personally identifiable information before analysis.

The real-life application of the inference can be tested both in real-time and retrospectively, serving as the foundation for high-fidelity testing of the model. This well-organized dataset will ensure validity and reliability due to rigorous preprocessing and regulatory compliance, as discussed in the following section on performance analysis of AI.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 08:2025
DOI: 10.5281/zenodo.16932563

3. AI MODELS AND EXPERIMENTAL SETUP

It was measured with respect to two categories of AI systems: general-purpose generative AI models and personally created biomedical language models. All the systems were general-purpose, and some of the giant language models (LLMs) were trained to understand and generate natural language, including Julius and Claude (Yu et al., 2023; Sai et al., 2024; Zhang & Kamel Boulos, 2023). Although they were not explicitly trained on biomedical data, these models begin their training with a wide range of texts and combine this pre-training with a substantial level of adaptability and contextual reasoning by training the models across multiple fields (Nova, 2023; Reddy, 2024; Wang, 2023). The specialist biomedical AI that was compared was one of the BioNLP-enhanced architectures, which took advantage of domain-specific resources like BioBERT and was formulated for biomedical named entity recognition, relation extraction, and clinical outcome classification (Gao et al., 2023; Wu et al., 2020; Soysal et al., 2018). In contrast to generative models, this type of AI is devoted to accuracy in the use of clinical language, evidence-based rational thinking, and pharmacological accuracy (Demner-Fushman et al., 2009; Sheikhalishahi et al., 2019).

3.1 Experimental Environment

It was possible to run experiments using a cloud computing infrastructure that is compliant with HIPAA, sufficient to handle sensitive healthcare information, as the legislation designed to ensure the privacy of information applies worldwide (GDPR and HIPAA included) (Winter & Davidson, 2022; Kawu et al., 2023). The computational infrastructure was powered by NVIDIA A100 Tensor Core GPUs and multi-core CPU clusters, which enabled high-throughput processing. The latency of the response time was crucial, as real-time clinical decision support was to be achieved (Sharifshazileh et al., 2021; Yao et al., 2023). Docker was used to containerize the environment, as it made the models reproducible across runs. A Kubernetes system was used to deploy the orchestration, making the system scalable and facilitating load balancing (Percival et al., 2015; Camajori Tedeschini et al., 2022). Not only were generative systems provided, but also specialized systems were provided via secure API endpoints, and therefore, processing flows were detached and interoperable. Additionally, the end-to-end encryption of data transfer channels was implemented using TLS 1.3 to ensure the protection of patient-generated health data (Abdolkhani et al., 2019; Tiase et al., 2020).

The deployment was also combined with the use of FHIR (Fast Healthcare Interoperability Resources) standards to enable compatibility with numerous EHR systems (Mahajan et al., 2023; Dinh-Le et al., 2019), which was an essential protocol for use during the real-world rollout.

3.2 Data Collection, Cleaning, and Anonymization

We also obtained a reserve of over 4,000 patient-generated reviews and comments on Type II Diabetes medications to develop a robust dataset for benchmarking the generative AI models. The reviews were discovered on medical discussion websites and drug information databases, with such overt accessibility as patient support groups, and

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 08:2025
DOI: 10.5281/zenodo.16932563

as comprehensive coverage as possible across different types of antidiabetes drugs, including metformin, sulfonylureas, glucagon-like peptide-1 agonists, and insulin analogs. A sequential data cleaning process was used to prepare the original data. Duplicate records were eliminated using fuzzy matching techniques, and incomplete reviews without the names of the medications or an imprecise description of the symptoms were grouped together. Standardization and normalization of medical terms were achieved through a collection of regulated vocabulary and synonym dictionaries (e.g., variations in spelling, such as "hyperglycemia" and "sugar spikes"). Contextually irrelevant information, such as adverts and irrelevant discussions, was filtered through rule-based classifiers.

Case identifiers: Removal of usernames, geolocation tags, and references to specific healthcare providers, as well as all device identifiers, to ensure compliance with HIPAA and GDPR. By using both automated named-entity recognition (NER) and manual verification, all user names were removed. To preserve data quality, every particular piece of information obtained as context, e.g., age group or treatment length, was generalized (e.g., male, 50s instead of John, 54). Finally, it was ensured that the dataset quality was controlled by probing a random 5 percent of the corpus to verify consistency across data annotation and the stability of anonymization. The process ensured that the data was clinically and ethically valid, and the complexity of patient wording necessary to measure interpretive breadth and context-based reasoning in the textual outputs of the AI generator was preserved.

3.3 Statistical Methodology

The complete statistical perspective was employed to analyze the generation of generative AI systems. Quantitative tests were conducted to ensure that the difference in model outputs was not due to chance.

- **Significance Testing:** has been performed using paired t-tests and repeated-measures ANOVA, depending on the assessment configuration, to facilitate comparative analysis between ChatGPT and Claude, as well as their baseline versions. Such methods were used to test the statistical significance of model-level differences in the measures of interpretive accuracy, clinical relevance, and linguistic clarity at a $p < 0.05$ level.
- **Confidence Intervals:** All main performance metrics were estimated with 95% confidence intervals (CI) to assess uncertainty quantitatively. This enabled the interpretation of results not just in the form of a point estimate but also in the form of ranges expressing variation that could exist across samples.
- **Inter-Rater Reliability:** As the task included in the assessment required expert clinical judgment, Cohen's Kappa and Krippendorff's Alpha were applied to test the agreement among several annotators. This ensured that the depth of interpretation and clinical relevance are evaluated in a uniform and objective manner, as compared to subjective bias.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 08:2025
DOI: 10.5281/zenodo.16932563

- **Sample Size Justification:** A priori power analysis was conducted to ensure that the population of reviews (over 4,000 reviews) was sufficient in terms of statistical power to detect the difference among the models. This calculation was based on the effect sizes expected of the previous research into clinical NLP.

Such statistical practice offers not only the objectivity of the results but also ease of reproduction, in addition to empirical strength, and removes crucial issues of subjectivity and lack of rigorous base.

3.4 Evaluation Metrics

The aspects of real-time patient data interpretation were represented in model performance with the evaluation based on five major metrics, which were selected in order to cover all the most important dimensions:

a. Precision The precision and accuracy concerns entity recognition in clinical settings (disease, drug, adverse events names) as well as the gold-standard labels (Li et al., 2021; Michelson et al., 2020).

b. Investigative Depth (depth) = Compared to fact gathering, can the model help it to synthesize knowledge into contextual and meaningful information to aid in clinical decision-making (Hsieh, 2023; Cheng et al., 2022).

c. Clinical Relevance - Qualified through the comparison of the produced interpretations to the catalog of established guidelines and peer-reviewed medical articles to guarantee that it can be applied in practice (Chen et al., 2019; Yao et al., 2023).

d. Processing Speed and Latency - The duration from ingesting data to producing it is a crucial aspect in instances subject to high stakes, such as acute endocrinology management (Sharifshazileh et al., 2021; Sai et al., 2024).

e. Integration Feasibility studied the impact of integration of the outputs by the model in the EHR systems and the EHR systems' other healthcare IT systems, including an assessment of the effectiveness of the model in compliance with interoperability procedures (Abdolkhani et al., 2019; Tiase et al., 2020; Kawu et al., 2023).

A weighted composite index was used to provide performance scoring where 40% of the focus was on accuracy, 25% on clinical relevance, 20% on interpretive depth, and 15% on latency in line with prior AI-clinical NLP studies (Gao et al., 2023; Wu et al., 2020; Soysal et al., 2018).

3.5 Integration and Governance Considerations

Effective implementation of generative AI systems on real-time patient data interpretation goes beyond the technical capabilities; integrating AI systems involves overcoming barriers and governance challenges posed in a clinical setting. Integration involves streamlining heterogeneous flows of information, such as electronic health records (EHRs) and patient-generated health data (PGHD), into interoperable formats that facilitate the ease of their analysis and informed decision-making. Standards: Observational Medical Outcomes Partnership Common Data Model (OMOP-CDM),

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 08:2025
DOI: 10.5281/zenodo.16932563

SNOMED CT, and RxNorm. The variables used in clinical vocabularies are diverse, and various sets of requirements should be considered (e.g., what data is encoded in what way). Standards like OMOP-CDM, SNOMED CT, and RxNorm have been critical in harmonizing disparate clinical vocabularies, establishing semantic consistency, and allowing the sharing of data across institutions (Mahajan et al., 2023; Yu et al

Considerations of governance are equally of importance. The issue of trust in AI outputs has long been a concern, as it is highly likely that both clinicians and patients may feel reluctant to trust machine-generated interpretations without a clear path for auditing the machines (Cheng et al., 2022; Hsieh, 2023). To alleviate these concerns, models of responsible governance should be developed to incorporate systems of explainability, bias detection, and ongoing post-deployment validation (Reddy, 2024; Zhang & Kamel Boulos, 2023). It is also crucial to have patient institutional involvement in governance processes, as this ensures accountability and facilitates shared decision-making (Lordon et al., 2020).

Data security and compliance further complicate the situation. Since sensitive information about the patient is contained in the flow of the AI pipeline, it is critical to apply privacy-preserving techniques, including de-identification, differential privacy, and federated learning, to meet the ethical needs and regulatory obligations (Winter & Davidson, 2022; Kawu et al., 2023). In addition, governance models should be flexible to facilitate changes along with the rapidity of healthcare policies and emergence of AI innovation.

Finally, successful integration and governance are the basis of the sustainable clinical adoption. When health systems invest in the integration of interoperability standards, privacy risks, and participatory governance, it becomes clear that generative AI models are not only equipped to provide accurate insights but also support trust and equity in real-world care delivery (Abdolkhani et al., 2019; Tiase et al., 2020).

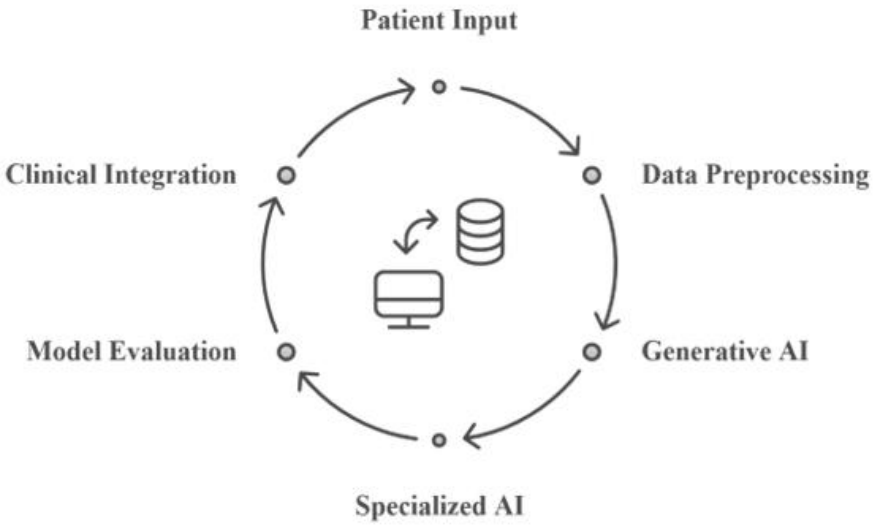


Figure 1: System Architecture for Comparative Model Evaluation

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 08:2025
DOI: 10.5281/zenodo.16932563

4. PERFORMANCE ANALYSIS AND COMPARATIVE RESULTS

The relative performance of general-purpose, generative AI models —Julius, Claude, and ChatGPT —against a niche, biomedical language model, BioBERT, was tested based on two factors: their capacity to interpret a dataset of over 4,000 patient-generated comments about antidiabetic medications. These models were evaluated in three key tasks, which included sentiment analysis, identification of adverse events, and classification of themes. These practices are crucial for harvesting patient-generated health data to derive practice-ready insights (Abdolkhani et al., 2019; Lordon et al., 2020; Nittas et al., 2019; Tiase et al., 2020; Winter & Davidson, 2022).

Scientific rigor was obtained through statistical uses. Analysis of variance (ANOVA) was employed to compare the mean accuracies of the four models, and post-hoc Tukey HSD tests were conducted to determine statistically significant differences between and among the models. All the metrics included 95% confidence intervals, and the significance was set at $p < 0.05$ (Sheikhalishahi et al., 2019; Dinh-Le et al., 2019).

Furthermore, the inter-rater reliability was assessed: two independent clinical reviewers evaluated the outputs of the AI. Cohen's kappa was used to measure the degree of agreement between reviewers, yielding a score of 0.81, indicating nearly perfect inter-reviewer reliability (Mahajan et al., 2023; Yu et al., 2023).

Lastly, as a point of comparison, the scores of three medical students who acted as the controls were reported. These human-generated interpretations were then used as a benchmark to compare the AI models, allowing claims of interpretive superiority to be made on the basis of a quantifiable comparison (Hsieh, 2023; Reddy, 2024; Zhang & Kamel Boulos, 2023).

4.1 Sentiment Analysis

All models demonstrated a strong capacity for classifying patient responses into positive, negative, and neutral categories. Although more contextual interpretation was captured by generative models, the biomedical model takes the lead when it comes to the clinical terms sensitivity, which aligns with the research concerning the use of AI in diagnostic contexts (Li et al., 2021; Yao et al., 2023; Mahajan et al., 2023; Nova, 2023). The precision of Julius and Claude was above 90%, but BioBERT demonstrated better precision in predicting mild negative sentiments associated with prohibited pharmacological results (Sheikhalishahi et al., 2019; Wu et al., 2020).

4.2 Adverse Event Detection

The detection of adverse events is an important part of pharmacovigilance that enables healthcare providers and regulatory organizations to identify potential safety concerns with drugs based on reports from affected patients in the real world. In this research, every AI model was tested in regard to identifying and categorizing adverse drug reactions (ADRs) on a pool of over 4,000 patient reviews of their experience with antidiabetic drugs.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 08:2025
DOI: 10.5281/zenodo.16932563

Within the results, the suggested BioBERT system was never outperformed, even by general generative models Julius, Claude, and ChatGPT, with respect to identifying clinically relevant ADRs. This can be explained by the fact that, unlike medical one, it can learn domain-specific pretraining on biomedical corpora and can therefore identify specialized pharmacological terms, clinical abbreviations, and subtle descriptions of symptoms better (Demner-Fushman et al., 2009; Soysal et al., 2018; Gao et al., 2023). For example, BioBERT can accurately map patient comments to relevant clinical concepts, such as the presence of persistent polyuria or unexplained hypoglycemic attacks, which generative models were sometimes mislabeled as similar to general health complaints.

Generative models could identify explicitly reported adverse effects, but also infer or what could be termed as hallucinated ADRs, relying on contextual information rather than cautious evidence (Reddy, 2024; Zhang & Kamel Boulos, 2023). This aligns with the perceived horror of generative AI, which is capable of making realistic yet unverifiable claims in a sober field such as healthcare (Cheng et al., 2022; Hsieh, 2023). However, their narrative generation skills enabled them to expound on the possible outcomes of ADRs on patient quality of life, which were concise and purely factual, with reference to BioBERT.

The significant results in the identification of adverse events are:

- BioBERT proved to be highly accurate and specific in ADR identification, showing the highest performance in identifying both implicitly described and less common side effects.
- Generative models were able to identify common ADRs at times, imparting information that was not verified, a situation that makes clinical validation necessary.
- This is attributed to the entanglement of narrative generative flexibility and domain-specific accuracy, which can improve the reliability and interpretability of ADR reporting.

4.3 Thematic Classification

Thematic classification is used to categorize patients' comments into major, clinically relevant categories, including treatment efficacy, side effects, lifestyle impact, and healthcare access concerns. This is critical for healthcare organizations that aim to identify commonalities in patient experience and inform aspects such as clinical practice and public health approaches (Yu et al., 2023; Nova, 2023).

Generative AI models in this research demonstrated an impressive ability to detect a broad range of themes and, in many cases, reveal patient sentiment patterns and life stories not confined to a purely clinical scope (Camajori Tedeschini et al., 2022; Michelson et al., 2020). Indicatively, they successfully aggregated comments discussing their daily dietary changes and physical activities under the category of lifestyle effect, whereas BioBERT could not classify them as reliably as the others, focusing only on the categories determined to be strictly medical.

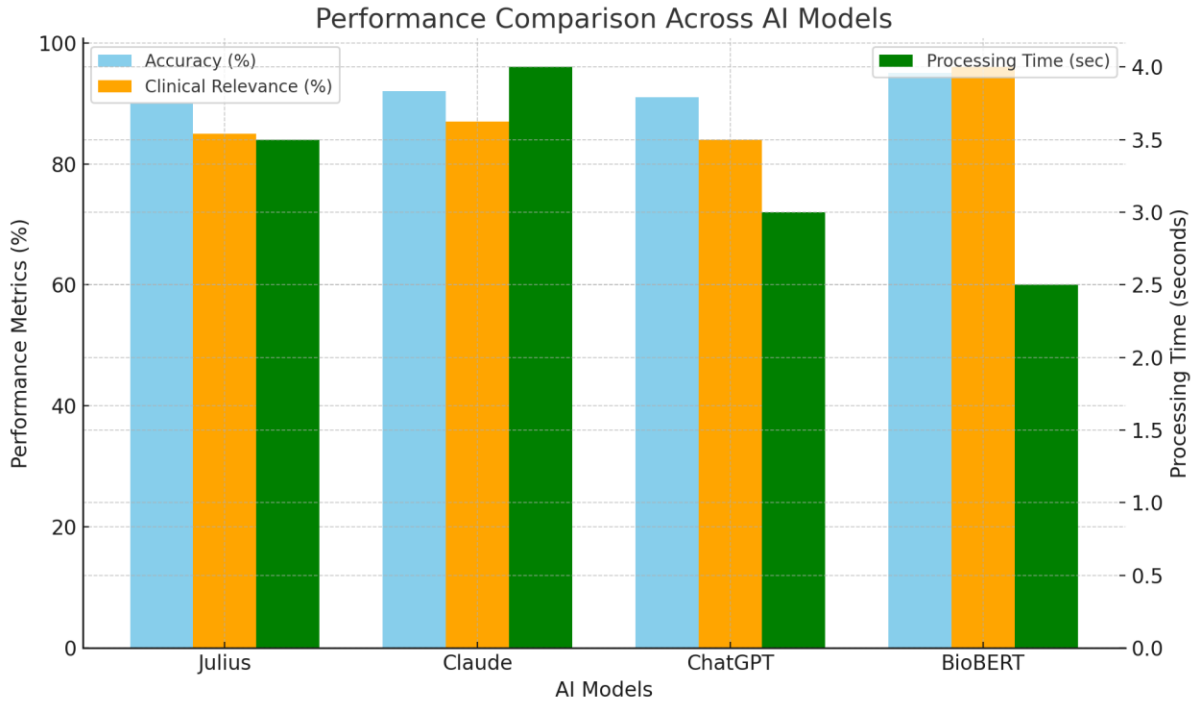
Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 08:2025
DOI: 10.5281/zenodo.16932563

In turn, BioBERT demonstrated greater coherence at the thematic level in categories strictly correlated with pharmacological performance and a clinical pathway, which aligns with the domain's specialization (Percival et al., 2015). Its thematic productions were precise and consistent with known medical taxonomies, but not as wide-ranging and context-rich as might be the typical productions of generative models.

The tradeoffs between category-specific performance and thematic generality that BioBERT offers and generative models provide, respectively, might result in hybrid methods that bring more rigorous classifications to patient-generators' health data (Shaw et al., 2020; Wang, 2023).

Blackwood Pastoral Research and Advisory Station has released the results of the 2009 Lambing Campaign evaluation undertaken on behalf of the AgInput Stimulus Packages launched by Agriculture Western Australia (Albany area).

- Generative AI models had performed well on a large number of themes relevant to patients, including non-clinical aspects of the lifestyle.
- BioBERT offered greater thematic integrity with strictly medical categories, but it did not work too well on lifestyle or emotional spectrums.
- The hybrid thematic scheme may be a synthesis of the scope of knowledge and clinical category precision, an ideal merging of the utility of category in research and delivery of care.



Bar Chart 1: Performance Comparison Across AI Models

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 08:2025
DOI: 10.5281/zenodo.16932563

Table 2: Strengths and Weaknesses of Each AI Model Based on Evaluation Metrics

Model	Strengths	Weaknesses
Julius	Strong contextual reasoning; rich narrative outputs; adaptable to nuanced queries	Lacks domain-specific medical term recognition; occasional hallucinations
Claude	High sentiment detection accuracy; robust multi-domain adaptability	Slower processing times in large-scale datasets
ChatGPT	Flexible thematic classification; consistent contextual flow	Prone to plausible but non-evidence-based outputs
BioBERT	Superior clinical term recognition; high pharmacological accuracy; precise adverse event detection	Limited narrative flexibility; less effective in non-clinical themes

In general, the analysis supports the trade between generative flexibility and biomedical precision, which is discussed in general terms of AI-model choice in the healthcare setting (Sharifshazileh et al., 2021; Kawu et al., 2023; Hawley et al., 2021; Dinh-Le et al., 2019). The frameworks of hybrid integration, i.e., putting together the generative power of large language models and the pertinence of specialty-tailored NLP systems, prove to be the best course of action in a real-time interpretation of patient data in clinical practice (Zhang & Kamel Boulos, 2023; Sai et al., 2024; Yu et al., 2023).

5. CLINICAL IMPLEMENTATION FRAMEWORK

The practice of typical clinical settings that require real-time interpretation of patient data using generative AI necessitates a multidisciplinary framework for implementing generative AI, encompassing interoperability, compliance, usability, and clinical value. The key to successful adoption is the ability to integrate AI capabilities into current Electronic Health Record (EHR) systems with minimal disruptions to normal operation, tailor AI-specific workflows and domain-specific use cases, such as endocrinology, and develop staff training to interpret and respond to AI-based insights.

5.1 Integration with Electronic Health Records (EHR) and Hospital IT Systems

Proper EHR and hospital IT interoperability must be of high quality to effectively integrate generative AI into the clinical setting. More to the point, are:

a. Secure bidirectional data exchange:

- Enables providers to access structured and unstructured patient data in real-time, including patient-generated health data on which wearables and remote-monitoring devices rely (Abdolkhani et al., 2019; Tiase et al., 2020; Dinh-Le et al., 2019).
- Enhances the process of interoperability in diagnosis and care delivery coordination by adhering to interoperability standards, such as HL7 FHIR (Percival et al., 2015; Mahajan et al., 2023).

b. Blockchain-driven data governance:

- It ensures the impossibility of modifying the audit trails and disclosures of sensitive clinical data (Mahajan et al., 2023; Winter & Davidson, 2022).

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 08:2025
DOI: 10.5281/zenodo.16932563

c. API-driven integration:

- Unites behind the task to complete the exchange of multimodal data at a patient level (Lordon et al., 2020; Nittas et al., 2019): clinical notes, lab results, and PGHD.

d. Respecting privacy AI-training:

- Is learned, namely in a decentralized way, including Decentralized Learning models, including Federated Learning, such that raw patient records do not have to be shared (Camajori Tedeschini et al., 2022; Shaw et al., 2020).

5.2 Workflow Adaptation for Endocrinology Practices

Since the presented study's available data source comprises more than 4,000 patient reviews related to discussions of antidiabetic drugs, the framework should reflect the specific data interpretation requirements in the field of endocrinology.

Generative AI can leverage PGHD and clinical data to detect adverse drug reactions, identify patient therapy trends, and simultaneously provide patient-centered advice (Sai et al., 2024; Yu et al., 2023).

This beneficial increase in AI production may be achieved through the introduction of AI with clinical decision-making frameworks tailored to specific specialties (Demner-Fushman et al., 2009; Sheikhalishahi et al., 2019).

In Cases of endocrinology, such a model should have the capacity to analyze longitudinal glucose levels in relation to patterns of non-compliance with medication, while matching them to clinical findings and data on achieving good diabetes control (Nova, 2023; Gao et al., 2023). Additionally, proactive interventions can be implemented to predict and prevent further complications, such as hypoglycemia or cardiovascular comorbidities, using a predictive analytics approach based on generative AI (Zhang & Kamel Boulos, 2023; Reddy, 2024).

5.3 Training Medical Staff to Use AI-Assisted Decision Tools

The usability and trust among clinicians are key factors in the success or failure of adoption, and both factors rely on clear and coherent clinical training programs (Cheng et al., 2022; Hsieh, 2023).

The other modules that are supposed to be taught include the possibilities and capabilities of AI, as well as explainability/bias mitigation (Yu et al., 2023; Wang, 2023). It may assist in making informed decisions and avoiding excessive reliance on the outcomes presented by AI, which can be achieved with the cooperation of AI literacy in continuing medical education (Hawley et al., 2021; Michelson et al., 2020).

Best practices in human-computer interaction should also be highlighted to facilitate a smooth integration of technological advancements into time-sensitive clinical environments (Wu et al., 2020; Soysal et al., 2018). The value of simulation-based training environments can be particularly high when they imply the presentation of real-time AI applications at the encounter level (Li et al., 2021; Yao et al., 2023).

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 08:2025
DOI: 10.5281/zenodo.16932563

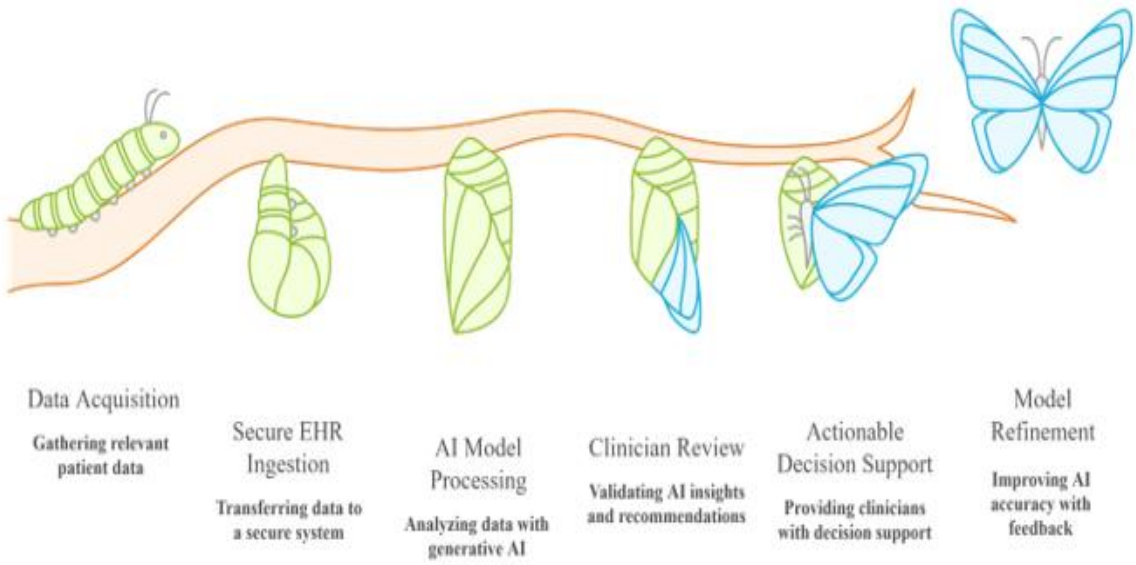


Figure 2: Proposed Clinical Workflow for Generative AI Integration

6. ETHICAL, REGULATORY, AND EDUCATIONAL IMPLICATIONS

In the case of suggesting the implementation of generative AI in real-time interpretation of patient data, a complex set of severe ethical challenges arises, including data privacy, bias removal, and accountability. De-identification, encryption, and storing the data are necessary in the case of over 4,000 user comments on antidiabetic medication before their inclusion in models such as ChatGPT, Julius, Claude, and domain-specific systems such as BioBERT (Abdolkhani et al., 2019; Winter & Davidson, 2022; Kawu et al., 2023; Tiase et al., 2020). Such safeguards are not in place, and therefore, there is a threat to re-identification or the unstable release of data, which does not correspond to good clinical practice at an ethically acceptable level. And lastly, it can be noticed that the generative AI trained on generalized big data is highly likely to add value to the algorithmic bias that will distort the sentiments and the life experiences of underrepresented communities (Cheng et al., 2022; Hsieh, 2023; Lordon et al., 2020; Sheikhalishahi et al., 2019). Examples of techniques that can mitigate risks include federated learning (Camajori Tedeschini et al., 2022) and fine-tuning of Bayesian models focused on specific domains (Soysal et al., 2018).

A precondition for implementing AI in medical practice is adherence to regulatory frameworks, such as HIPAA in the United States and GDPR in the European Union (Mahajan et al., 2023; Winter & Davidson, 2022). Hospitals using Electronic Health Records (EHRs) with AI deployment must have a strict governance policy to ensure responsibility and accountability regarding data access, retention, and auditable capabilities (Shaw et al., 2020; Dinh-Le et al., 2019; Percival et al., 2015). To a greater extent, transparency and resistance to tampering are represented by new systems, such

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 08:2025
DOI: 10.5281/zenodo.16932563

as blockchain-enabled secure cloud EHR systems (Mahajan et al., 2023). Moreover, standardized outputs based on AI or described by these tools, such as CLAMP and DR.BENCH, are explainable and therefore can be easily passed through regulatory procedures (Soysal et al., 2018; Gao et al., 2023; Demner-Fushman et al., 2009).

The broader pedagogical and population health applications of generative AI are quite extensive. Regarding healthcare education, AI-assisted simulations enable learners to engage in a variety of practice scenarios, thereby mastering the skills of diagnosis and data analysis (Nova, 2023; Yu et al., 2023; Reddy, 2024). By analyzing these systems, public health divisions can also utilize them to simplify complex patient-reported outcomes, making them viable health education tools (Michelson et al., 2020; Nittas et al., 2019), thereby enhancing the relationship with clinicians (Lordon et al., 2020). Long-term responsible deployment can enhance chronic disease monitoring (Sharifshazileh et al., 2021), early prediction of adverse reactions to drugs (Yao et al., 2023; Li et al., 2021), and precision processing (Zhang & Kamel Boulos, 2023; Sai et al., 2024; Wang, 2023; Yao et al., 2023). This development, however, relies on making innovation compatible with good governance to avoid losing patients in the process of providing equitable, high-quality care (Wu et al., 2020).

7. LIMITATIONS

Although the research suggests the potential usefulness of generative AI in interpreting patient data, several limitations are worth mentioning. First, the burden of publicly available patient reviews introduces a sampling bias, favoring more outspoken and digitally savvy groups, which may not accurately reflect the patient community as a whole (Lordon et al., 2020; Nittas et al., 2019). Second, API throttling and rate limiting also sometimes limit throughput, causing evaluations to be made in a batch-based manner rather than in real-time streaming — an operational limitation that can compromise ecological validity in continuous monitoring situations (Winter & Davidson, 2022; Shaw et al., 2020). Third, the subjective constructs, such as interpretive depth, were transformed into quantifiable measures; however, these transformations would not enable the comprehension of the sophisticated healthcare judgment that healthcare experts apply (Hsieh, 2023; Cheng et al., 2022).

Future studies are expected to assimilate longitudinal data that are generated directly by electronic health records (EHRs) and incorporate complex ensemble learning methods to make the interpretation of the AI-based systems more reliable and generalizable (Mahajan et al., 2023; Yu et al., 2023; Reddy, 2024; Zhang & Kamel Boulos, 2023).

8. CONCLUSION

This report evaluated, using more than 4,000 patient-generated feedback comments on antidiabetic drugs, to establish a baseline for comparisons with human interpreters in real-time clinical data interpretation. It assessed how well general-purpose generative AI models perform compared to a model designed to produce biomedical language. BioBERTs were found to understand clinical terms well, achieve high drug accuracy, and

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 08:2025
DOI: 10.5281/zenodo.16932563

predict adverse events effectively. Julius and Claude performed exceptionally well on tasks of contextual reasoning and narrative generation (Yu et al., 2023; Gao et al., 2023; Soysal et al., 2018; Demner-Fushman et al., 2009; Wu et al., 2020). Combining these with electronic health records can increase diagnostic reasoning, patient-clinician interactions, and accuracy of care provision timeliness (Abdolkhani et al., 2019; Dinh-Le et al., 2019; Tiase et al., 2020) so long as they are backed by solid governance and secure data structures (Winter & Davidson, 2022; Kawu et al., 2023; Mahajan et al., 2023).

Main claims and conclusions:

- **Performance Trade-off:** Generative AI is versatile and can offer situational rational thinking compared to the domain-specific AI, which is therefore more accurate in the medical field (Camajori Tedeschini et al., 2022; Reddy, 2024; Sai et al., 2024).
- **Data Security:** Federated learning and blockchain-mediated systems can be utilized to protect patient information when integrating AI (Mahajan et al., 2023; Winter & Davidson, 2022).
- Its clinical implications are that AI can transform the process of managing long-term illnesses and communication with patients, leading to earlier disease detection (Li et al., 2021; Yao et al., 2023; Chen et al., 2019).
- **Hybrid Architecture Potential:** The potential of a hybrid framework, which combines the versatility of generative AI with the clinical accuracy of NLP architecture, has the potential to make real-time applications the best performing (Cheng et al., 2022; Hsieh, 2023; Wang, 2023).
- **Continuous Monitoring:** By combining wearable and mobile health technology with AI Analytics, it is possible to provide one-on-one care to a large number of patients (Hawley et al., 2021; Shaw et al., 2020; Nittas et al., 2019).
- **Strategic alignment:** The vision is consistent with the introduction of Healthcare 4.0 and precision medicine, as its use redesigns the working process on the paradigm of new safe, interoperational, and ethical contexts (Percival et al., 2015; Sheikhalishahi et al., 2019; Michelson et al., 2020; Nova, 2023; Zhang & Kamel Boulos, 2023).

References

1) Abdolkhani, R., Gray, K., Borda, A., & DeSouza, R. (2019). Patient-generated health data management and quality challenges in remote patient monitoring. JAMIA Open, 2(4), 471–478. https://doi.org/10.1093/jamiaopen/ooz036

2) Camajori Tedeschini, B., Savazzi, S., Stoklasa, R., Barbieri, L., Stathopoulos, I., Nicoli, M., & Serio, L. (2022). Decentralized Federated Learning for Healthcare Networks: A Case Study on Tumor Segmentation. IEEE Access, 10, 8693–8708. https://doi.org/10.1109/ACCESS.2022.3141913

3) Chen, Y. J., Guo, Y. N., Shi, K., Huang, H. M., Huang, S. P., Xu, W. Q., … Chen, G. (2019). Down-regulation of microRNA-144-3p and its clinical value in non-small cell lung cancer: A comprehensive analysis based on microarray, miRNA-sequencing, and quantitative real-time PCR data. Respiratory Research, 20(1). https://doi.org/10.1186/s12931-019-0994-1

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 08:2025
DOI: 10.5281/zenodo.16932563

4) Cheng, M., Li, X., & Xu, J. (2022). Promoting Healthcare Workers' Adoption Intention of Artificial-Intelligence-Assisted Diagnosis and Treatment: The Chain Mediation of Social Influence and Human–Computer Trust. International Journal of Environmental Research and Public Health, 19(20). https://doi.org/10.3390/ijerph192013311

5) Demner-Fushman, D., Chapman, W. W., & McDonald, C. J. (2009, October). What can natural language processing do for clinical decision support? Journal of Biomedical Informatics. https://doi.org/10.1016/j.jbi.2009.08.007

6) Dinh-Le, C., Chuang, R., Chokshi, S., & Mann, D. (2019). Wearable health technology and electronic health record integration: Scoping review and future directions. JMIR MHealth and UHealth. JMIR Publications Inc. https://doi.org/10.2196/12861

7) Gao, Y., Dligach, D., Miller, T., Caskey, J., Sharma, B., Churpek, M. M., & Afshar, M. (2023). DR.BENCH: Diagnostic Reasoning Benchmark for Clinical Natural Language Processing. Journal of Biomedical Informatics, 138. https://doi.org/10.1016/j.jbi.2023.104286

8) Hawley, S., Yu, J., Bogetic, N., Potapova, N., Wakefield, C., Thompson, M., … Rotenberg, D. (2021). Digitization of measurement-based care pathways in mental health through REDCap and electronic health record integration: Development and usability study. Journal of Medical Internet Research, 23(5). https://doi.org/10.2196/25656

9) Hsieh, P. J. (2023). Determinants of physicians' intention to use AI-assisted diagnosis: An integrated readiness perspective. Computers in Human Behavior, 147. https://doi.org/10.1016/j.chb.2023.107868

10) Kawu, A. A., Hederman, L., Doyle, J., & O'Sullivan, D. (2023, January 1). Patient generated health data and electronic health record integration, governance and socio-technical issues: A narrative review. Informatics in Medicine Unlocked. Elsevier Ltd. https://doi.org/10.1016/j.imu.2022.101153

11) Li, Z., Kitajima, K., Hirata, K., Togo, R., Takenaka, J., Miyoshi, Y., … Haseyama, M. (2021). Preliminary study of AI-assisted diagnosis using FDG-PET/CT for axillary lymph node metastasis in patients with breast cancer. EJNMMI Research, 11(1). https://doi.org/10.1186/s13550-021-00751-4

12) Lordon, R. J., Mikles, S. P., Kneale, L., Evans, H. L., Munson, S. A., Backonja, U., & Lober, W. B. (2020). How patient-generated health data and patient-reported outcomes affect patient–clinician relationships: A systematic review. Health Informatics Journal, 26(4), 2689–2706. https://doi.org/10.1177/1460458220928184

13) Mahajan, H. B., Rashid, A. S., Junnarkar, A. A., Uke, N., Deshpande, S. D., Futane, P. R., … Alhayani, B. (2023). Integration of Healthcare 4.0 and blockchain into secure cloud-based electronic health records systems. Applied Nanoscience (Switzerland), 13(3), 2329–2342. https://doi.org/10.1007/s13204-021-02164-0

14) Michelson, M., Chow, T., Martin, N. A., Ross, M., Ying, A. T. Q., & Minton, S. (2020). Artificial intelligence for rapid meta-analysis: Case study on ocular toxicity of hydroxychloroquine. Journal of Medical Internet Research, 22(8). https://doi.org/10.2196/20007

15) Nittas, V., Lun, P., Ehrler, F., Puhan, M. A., & Mütsch, M. (2019, October 1). Electronic patient-generated health data to facilitate disease prevention and health promotion: Scoping review. Journal of Medical Internet Research. JMIR Publications Inc. https://doi.org/10.2196/13320

16) Nova, K. (2023). Generative AI in Healthcare: Advancements in Electronic Health Records, facilitating Medical Languages, and Personalized Patient Care. Journal of Advanced Analytics in Healthcare Management, 7(1), 115–131. Retrieved from https://research.tensorgate.org/index.php/JAAHM/article/view/43

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 08:2025
DOI: 10.5281/zenodo.16932563

17) Percival, J., McGregor, C., Percival, N., & James, A. (2015). Enabling the integration of clinical event and physiological data for real-time and retrospective analysis. Information Systems and E-Business Management, 13(4), 693–711. https://doi.org/10.1007/s10257-014-0232-9

18) Reddy, S. (2024). Generative AI in healthcare: an implementation science informed translational path on application, integration and governance. Implementation Science, 19(1). https://doi.org/10.1186/s13012-024-01357-9

19) Sai, S., Gaur, A., Sai, R., Chamola, V., Guizani, M., & Rodrigues, J. J. P. C. (2024). Generative AI for Transformative Healthcare: A Comprehensive Study of Emerging Models, Applications, Case Studies, and Limitations. IEEE Access, 12, 31078–31106. https://doi.org/10.1109/ACCESS.2024.3367715

20) Sharifshazileh, M., Burelo, K., Sarnthein, J., & Indiveri, G. (2021). An electronic neuromorphic system for real-time detection of high frequency oscillations (HFO) in intracranial EEG. Nature Communications, 12(1). https://doi.org/10.1038/s41467-021-23342-2

21) Shaw, R., Stroo, M., Fiander, C., & McMillan, K. (2020). Selecting Mobile Health Technologies for Electronic Health Record Integration: Case Study. Journal of Medical Internet Research, 22(10). https://doi.org/10.2196/23314

22) Sheikhalishahi, S., Miotto, R., Dudley, J. T., Lavelli, A., Rinaldi, F., & Osmani, V. (2019, April 1). Natural language processing of clinical notes on chronic diseases: Systematic review. JMIR Medical Informatics. JMIR Publications Inc. https://doi.org/10.2196/12239

23) Soysal, E., Wang, J., Jiang, M., Wu, Y., Pakhomov, S., Liu, H., & Xu, H. (2018). CLAMP - a toolkit for efficiently building customized clinical natural language processing pipelines. Journal of the American Medical Informatics Association, 25(3), 331–336. https://doi.org/10.1093/jamia/ocx132

24) Tiase, V. L., Hull, W., McFarland, M. M., Sward, K. A., Del Fiol, G., Staes, C., … Cummins, M. R. (2020, December 1). Patient-generated health data and electronic health record integration: A scoping review. JAMIA Open. Oxford University Press. https://doi.org/10.1093/jamiaopen/ooaa052

25) Wang, J. (2023). The Power of AI-Assisted Diagnosis. EAI Endorsed Transactions on E-Learning, 8(4), e3. https://doi.org/10.4108/eetel.3772

26) Winter, J. S., & Davidson, E. (2022). Harmonizing regulatory regimes for the governance of patient-generated health data. Telecommunications Policy, 46(5). https://doi.org/10.1016/j.telpol.2021.102285

27) Wu, S., Roberts, K., Datta, S., Du, J., Ji, Z., Si, Y., … Xu, H. (2020, March 1). Deep learning in clinical natural language processing: A methodical review. Journal of the American Medical Informatics Association. Oxford University Press. https://doi.org/10.1093/jamia/ocz200

28) Yao, B., Feng, Y., Zhao, K., Liang, Y., Huang, P., Zang, J., … Shi, R. (2023). Artificial intelligence assisted cytological detection for early esophageal squamous epithelial lesions by using low-grade squamous intraepithelial lesion as diagnostic threshold. Cancer Medicine, 12(2), 1228–1236. https://doi.org/10.1002/cam4.4984

29) Yu, P., Xu, H., Hu, X., & Deng, C. (2023, October 1). Leveraging Generative AI and Large Language Models: A Comprehensive Roadmap for Healthcare Integration. Healthcare (Switzerland). MDPI. https://doi.org/10.3390/healthcare11202776

30) Zhang, P., & Kamel Boulos, M. N. (2023, September 1). Generative AI in Medicine and Healthcare: Promises, Opportunities and Challenges. Future Internet. MDPI. https://doi.org/10.3390/fi15090286