

Computational Use of Patient–Provider Secure Messaging Data to Achieve Better Clinical Efficiency and Quality of Communication: A Systematic Review

Yawen Guo, MISM¹, Di Hu, MS¹, Yiliang Zhou, MS¹, Tianchu Lyu, PhD¹, Kai Zheng, PhD¹
¹ University of California Irvine, Irvine, CA, United States

Abstract

Secure messaging (SM) between patients and providers has seen increasing adoption over the past decade, prompting development of computational methods to enable automation and research to enhance clinical efficiency and quality of communication. Through a systematic review, we examined the extant literature to investigate how previous studies had applied computational analyses to SM data. After screening 1,374 papers, we identified 19 relevant studies published between 2017 and 2024, most of which focused on applications for streamlining clinical workflows, facilitating early disease detection, supporting personalized decision making, and enhancing patient health literacy. Among the computational methods used, BERT was consistently shown to deliver best performance. However, all existing studies were constrained by small-size datasets and limited healthcare settings, leading to inadequate validation and poor generalizability. The results of this review highlight key research gaps, particularly the need for more robust computational approaches that ensure scalability, fairness, and clinical applicability.

Keywords. secure messaging, patient–provider communication, computational methods, systematic review

Introduction

The Health Information Technology for Economic and Clinical Health (HITECH) Act was enacted in 2009 to promote the meaningful use of health information technology by healthcare providers¹. Meaningful use refers to the utilization of certified electronic health records (EHR) to improve the quality of care, engage patients, and ensure the security of patient information. It includes a specific requirement for eligible providers to use secure electronic messaging to communicate with patients to extend access to care, improve transparency, and safeguard patient confidentiality².

Secure messaging (SM), as defined by the Centers for Medicare and Medicaid Services, refers to “any electronic communication between a provider and patient that ensures only those parties can access the communication”³. It has become a widely adopted feature in EHR systems and has supported various functionalities such as appointment scheduling, medication refill, lab notification, and delivery of web-based interventions^{4,5}. By facilitating communication between patients and providers, SM has become an indispensable tool in modern healthcare organizations to enhance patient empowerment and satisfaction. However, overuse of SM has also led to a growing concern, particularly during and after the COVID-19 pandemic, for its role in increasing the burden on providers, contributing to their burn-out^{6–8}. To improve the efficiency in processing large volumes of patient-initiated secure messages, artificial intelligence (AI)-based tools have been developed to help manage message triage and response, accompanied by the introduction of new billing policies for e-visits and secure messaging^{9–11}.

Previous reviews on SM have explored its integration into healthcare systems and its impact on communication efficiency and quality. For example, reviews conducted by McGeedy et al., Wallwiener et al., Goldzweig et al.—and a few others focusing on specific disease topics—have highlighted the evolution of electronic communication tools, assessing their benefits such as improved healthcare delivery and patient satisfaction, while also addressing challenges such as confidentiality and legal issues^{12–15}. A comprehensive understanding of the research literature on SM is crucial for identifying knowledge gaps and informing the development of future tools to enhance clinical workflow, reduce clinician burden, and improve patient care^{16–18}.

While there have been studies devoted to applying computational methods to real-world SM data, no systematic review exists that synthesizes these efforts to provide a better understanding on how they contribute to advancing patient–provider communication, clinical decision making, and clinical and health services research. The systematic review study reported in this paper addresses this gap by examining existing models and tools developed from real-world SM data. During the review, we focused on classifying the study objectives, methodologies, data sources and scope, evaluation metrics, and clinical applications of existing research; in addition to providing a structured

synthesis of applications built on real-world SM data and identifying key trends, limitations, and directions for future research.

Methods

We conducted a systematic literature search across databases including PubMed, Scopus, IEEE Xplore, ACM Digital Library, Cochrane CENTRAL, CINAHL, and Web of Science. The search strategy used the following terms: ("secure messag*" AND patient) OR ("portal messag*" AND patient) in the title and abstract, restricted to studies published from 2009 onwards. Only articles written in English were included. To ensure the robustness of the search strategy, all searches were repeated and reviewed by a librarian for accuracy. The final search was completed in October, 2024.

We included studies that leveraged real-world SM data between patients and providers to develop computational applications and tools. Since this review study focused on patient–provider communication, studies were excluded if they did not use real-world SM content or if the SM did not occur between patients and providers—for example, communication among caregivers, medical students and faculty members, or across different medical settings (e.g., care transitions). We also excluded studies that examined SM alongside other communication methods and reported results collectively. Additionally, studies were excluded if they solely used existing methods or algorithms for evaluation, without developing new applications or tools based on real-world SM data. We did not include opinion pieces, media articles, commentaries, review articles, abstracts, posters, brief communications, or research letters. The titles and abstracts of the identified articles were independently screened by two authors (YG and DH) using the inclusion and exclusion criteria. Any disagreements were resolved through consensus meetings with a third author (KZ). Articles identified as relevant after the title and abstract screening process then underwent a full-text review to determine final inclusion.

The data extracted from the included papers followed a predetermined structure to systematically capture key elements relevant to our study objective. For computational models and applications developed, we extracted information on the algorithm's general category (e.g., classification, topic modeling, prediction, named entity recognition), study objective, data source and scope, primary evaluation metrics and performance, comparative evaluation, results, outcomes, and recommendations for future research. Each article was independently coded by two of three researchers (YG, DH, and YZ), all of whom are actively conducting research in clinical informatics and patient–provider digital communication. Discrepancies between reviewers were resolved during consensus meetings or with the input of the third researcher (KZ) as needed.

Results

We identified 2,151 studies from the databases using the search strategy outlined above. A total of 1,374 duplicate references were identified using Covidence software, and an additional 29 were removed manually. We then screened the titles and abstracts of the remaining 748 studies, excluding 358 that did not meet the inclusion criteria, with full text screening leading to the removal of an additional 391 references. A total of 19 studies were included in the final review. Figure 1 provides an overview of the literature search and screening process.

We identified 19 studies focusing on the computational use of patient–provider SM data to develop various applications and tools, published between 2017 and 2024. Most of these studies were published in leading medical informatics venues including Journal of the American Medical Informatics Association (JAMIA) and Journal of Medical Internet Research (JMIR). The studies developed a diverse range of computational approaches to address challenges in patient–provider SM communication through patient portal. Text classification models were the most prevalent ($n = 15$, 78.9.3%), utilizing methods ranging from traditional ML models, such as logistic regression (LR), support vector machines (SVM), and random forest (RF), to deep learning (DL) models, particularly BERT and its variants^{19–22}. Ensemble approaches integrating multiple natural language processing (NLP) techniques were also employed to improve classification accuracy and contextual understanding, such as fusion models combining convolutional neural networks (CNNs) with transformer-based architectures^{22,23}. Beyond classification tasks, the studies explored predictive modeling, unsupervised learning, and word embedding optimization to develop computational applications for SM data. One study introduced an automated readability formula using a novel NLP-based approach to predict text difficulty, outperforming traditional Flesch-Kincaid readability scores in assessing message comprehensibility²⁴. A study on health literacy profiling developed and validated computational

linguistic models to assess patient literacy levels, analyzing 283,216 SMs from diabetes patients to demonstrate that automated literacy profiles could outperform traditional health literacy assessments²⁵. A tailored word embedding optimization model called PK-word2vec was designed to address the challenges of small SM datasets, aiming to enhance NLP applications in clinical settings by improving domain-specific language representation²⁶. Additionally, a large-scale study leveraging a dataset of two million SMs applied unsupervised learning to uncover common patient concerns and communication patterns, enabling a data-driven exploration of themes in patient-provider interactions²⁷.

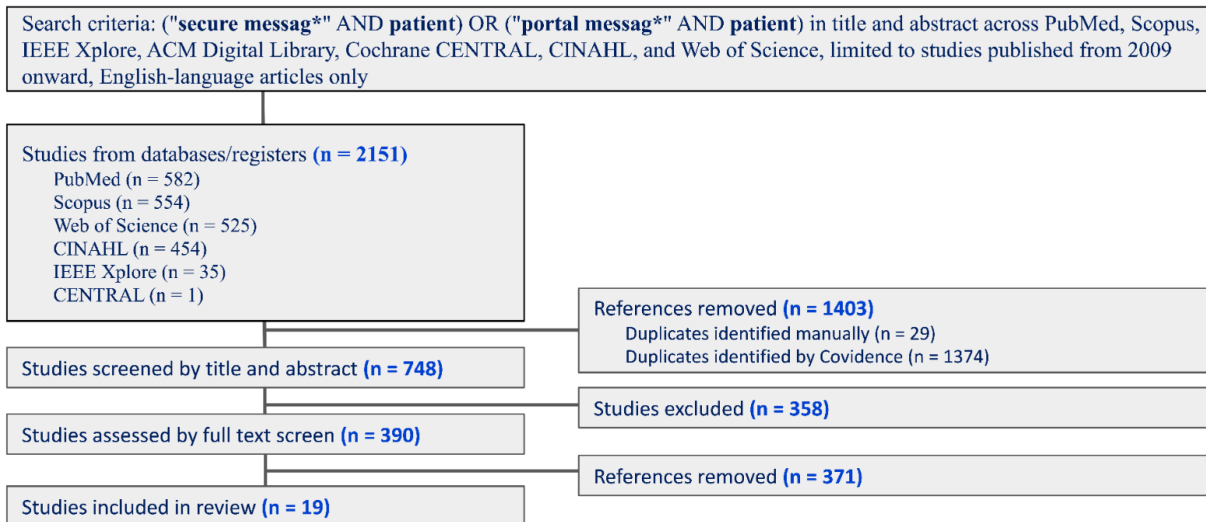


Figure 1. Search strategy and PRISMA flow diagram for the selection of literature

Study Objectives. These computational applications were primarily designed to achieve three major objectives: optimizing clinical workflow and triage processes, reducing clinician burden, and enhancing patient communication and health literacy.

Optimizing Clinical Workflow and Message Triage (n=5, 26.3%). Studies aimed to streamline message triage by automatically prioritizing patient messages and classifying urgent and non-urgent inquiries. Text classification techniques were leveraged to categorize messages based on urgency, clinical relevance, and real-time prioritization^{20,21,28}. One study specifically classified patient-initiated EHR messages for COVID-19 triage related to medication requests (e.g., Paxlovid), aiming to reduce response time and improve access to antiviral treatment²⁹. Additionally, NLP-based approaches were used to predict message complexity and medical decision-making difficulty, further improving workflow efficiency in clinical settings³⁰.

Reducing Clinician Burden (n=8, 42.1%). Several studies sought to alleviate clinician workload by automating repetitive tasks and facilitating efficient decision-making. Machine learning applications were developed to detect hypoglycemia incidents in SMs³¹, depression concerns¹⁹, and identify proxy-authored messages in oncology settings³². Several studies classified patient-initiated EHR messages using similar patterns, grouping them into symptoms, prescriptions, logistical concerns, and general updates³³⁻³⁵. Some studies further distinguished between informational, medical, and social needs^{33,36}. Despite variations in terminology, these studies shared a focus on organizing messages to improve response efficiency and reduce clinician burden.

Enhancing Patient Communication and Health Literacy (n=6, 31.6%). Another major objective was improving patient-provider interactions and understanding patient needs. Several studies developed NLP-based classifiers to assess patient health literacy and predict the readability of physician messages^{24,25,37,38}. Others leveraged topic modeling and named entity recognition to extract key themes from patient messages to better understand patient needs³⁹, with one study specifically uncovering social needs such as food and housing insecurity, and transportation needs to ensure timely referrals and interventions³⁶.

Data Scale and Source. The scale of data used in the studies varied significantly. We used a cutoff of 10,000 messages to distinguish between small-scale and large-scale studies. The small-scale studies (n=13, 68.4%) primarily focused on analyzing patient-generated messages within limited clinical contexts, such as decision complexity, communicative health literacy, and provider-patient communication patterns^{19,21,23,24,28,30–34,36,38,40}. Smaller datasets allowed for greater depth in qualitative analyses but limited interoperability for broader implementation. Large-scale studies (n=6, 31.6%) incorporated datasets of 10,000 or more messages^{20,22,25,26,37,39}. These large-scale approaches enabled the application of sophisticated computational approaches but required automated annotation pipelines. The variation in data scale influenced the methodologies used across studies. Small-scale studies relied more on manual annotations, expert reviews, and traditional ML approaches. Large-scale studies leveraged deep learning architectures, such as BERT-based models, and required automated processing pipelines to handle the volume of data effectively.

Best Performing Models. Across studies, deep learning-based models, particularly BERT and its variants, consistently outperformed traditional ML approaches in classification and triage tasks^{19–21,23,29}. For example, a BERT-based large language model achieved an area under the receiver operating characteristic curve (AUROC) of 0.97 in prioritizing patient messages, demonstrating its potential for clinical deployment⁴¹. Traditional approaches like logistic regression, support vector machine, and random forest, while less dominant, remained competitive in tasks requiring structured feature extraction such as detecting mental health crises and social needs concerns^{31,33,40}.

Comparison of Model Performance. Studies employed various evaluation metrics based on specific tasks. For message triage, prioritization, and information extraction tasks such as extracting patient social needs, metrics such as AUROC, Precision, Recall, and F1-score were commonly used to assess how well models distinguished between urgent and non-urgent messages^{21,23,28–30,33,35}. Health literacy and readability analysis utilized readability scores and human ratings to evaluate the effectiveness of NLP models in predicting patient understandability^{20,24,25,37,38}. Clinical decision complexity and mental health-related studies relied on classification accuracy, inter-rater agreement, sensitivity, and specificity, ensuring models could reliably distinguish between different levels of clinical significance^{19,30,31}. The diversity in evaluation metrics reveals that model evaluation must be tailored to the specific clinical applications, balancing precision, recall, and interpretability to improve real world implementation in patient care.

Table 1. Study Characteristics

Study ID	Model Category	Study Objective	Data Source and Scope
Crossley 2020 ²⁴	Readability Prediction (NLP, Linguistic Feature Extraction)	Develop an automated readability formula to predict human ratings of physician text difficulty in SMs.	724 messages from physicians and patients from Kaiser Permanente Northern California(KPNC).
Schillinger 2021 ²⁵	Health Literacy Prediction (NLP & ML)	Develop automated literacy profiles for identifying patients with limited health literacy using computational linguistics.	283,216 SMs from 6,941 diabetes patients in the KPNC system
Yang 2024 ²⁰	Text Classification (BERT-based large language model(LM))	Develop and evaluate an AI system to prioritize patient portal messages for rapid review by clinicians.	40,132 patient-sent messages from NYU Langone Health; 7,260 expert-reviewed for validation.
Si 2020 ²¹	Text Classification (BERT-based model with Label Embeddings and Knowledge Distillation)	Develop LESA-BERT, a BERT-based model with label embeddings and knowledge distillation for improved classification of patient message urgency.	1,756 patient messages from a university hospital portal, categorized as non-urgent, medium, or urgent by clinicians.
Mermin-Bunnell 2023 ²⁹	Text Classification (Bidirectional Encoder Representations from Transformers - BERT)	Develop a model to classify patient-initiated messages for COVID-19 triage, reducing response time and improving access to antiviral treatment.	10,172 messages categorized into COVID-19 positive, COVID-19 other, and non-COVID-19 messages.

Sulieman 2017 ⁴⁰	Text Classification (Convolutional Neural Networks - CNN)	Assess the effectiveness of CNNs for classifying messages into informational, medical, social, and logistical.	3,000 messages from the My Health At Vanderbilt portal, manually classified into four categories.
Ren 2024 ²³	Text Classification (Fusion framework of pretrained LMs)	Enhance automated identification of patient primary concerns in portal messages.	2,239 patient portal messages from Mayo Clinic categorized into 4 primary concern classes.
Benda 2022 ³²	Text Classification (Least Absolute Shrinkage and LASSO)	Develop an NLP-based model to identify nonpatient authors (unregistered proxies) of patient portal messages in oncology.	2,000 patient portal messages from MSKCC, manually labeled as patient-authored or proxy-authored.
Cronin 2017 ²⁸	Text Classification (LR, Naïve Bayes, RF)	Compare rule-based and ML methods for classifying patient portal messages into different communication types.	3,253 patient portal messages from Vanderbilt University Medical Center (2005-2014).
Buchem 2024 ¹⁹	Text Classification (LR, SVM, BERT, RedditBERT)	Develop NLP-based classifiers to detect depression concerns in patient messages from cancer patients using traditional ML and deep learning approaches.	6600 patient messages from a comprehensive cancer cohort at Stanford Cancer Center, annotated by healthcare professionals.
Chen 2019 ³¹	Text Classification (LR, SVM, RF)	Develop HypoDetect, an NLP-based system, to detect hypoglycemia incidents reported by patients in SMS to facilitate early clinical response.	3,000 message threads from the US Department of Veterans Affairs, annotated for hypoglycemia incidents with 3.80% positive cases.
Sulieman 2020 ³⁰	Text Classification (Multinomial Naïve Bayes, RF)	Develop and evaluate ML classifiers to quantify and classify the complexity of medical decision-making in patient portal messages.	500 message threads from surgical patients at Vanderbilt University Medical Center, manually labeled for decision complexity.
Cronin 2019 ³³	Text Classification (Naïve Bayes, LR, RF)	Develop automated classifiers using NLP and ML techniques to identify consumer health information needs expressed in patient portal messages.	1,000 patient-generated messages from Vanderbilt University Medical Center's My Health at Vanderbilt (MHAV) patient portal.
Balyan 2019 ³⁷	Text Classification (NLP and ML)	Develop automated literacy profiles to classify health literacy levels from secure patient messages using NLP and ML.	283,216 messages from KPNC, validated against self-reported health literacy.
Crossley 2021 ³⁸	Text Classification (NLP on Linguistic Features)	Develop automatic models of patient communicative health literacy using linguistic features extracted from SMS.	9,530 patient SMS from KPNC, with a subset of 512 messages annotated for communicative health literacy.
Moon 2024 ⁴²	Text Classification (NLP sentence similarity and rule-based methods)	Develop NLP methods to identify unmet social needs in clinical text for timely intervention and CHW referrals.	5,675 clinical notes and 475 portal messages from 1,103 patients referred to a CHW program.
A 2019 ³⁵	Text Classification (Named Entity Recognition(NER), Ensemble DL)	Develop an ensemble DL model combined with rule-based NER to organize messages into active symptoms and logistical requests.	6000 patient portal messages from Mayo Clinic (2018-2019), labeled as Active Symptom or Logistical.

De 2021 ³⁹	Topic Modeling (Latent Dirichlet Allocation)	Develop a FHIR-based data model to analyze and extract significant information from messages.	More than 2 million patient-generated SMs from Mayo Clinic.
Song 2024 ²⁶	Word Embedding Optimization	Develop PK-word2vec, an adapted word embedding model for small datasets to improve semantic representation.	137,554 messages from 1,389 breast cancer patients at Vanderbilt University Medical Center.

Discussion

Previous studies on patient portals and SM systems have examined system usability, SM communication efficiency, and provider workload. While early evaluations highlighted responsiveness of these systems⁴³, later analyses identified barriers such as limited message access, delays in response times⁴⁴, and burdens on clinicians⁴⁵. The COVID-19 quarantine policy and the recent shift in e-visit billing policies have significantly influenced the SM usage and communication patterns, raising new questions about efficiency and economic implications^{9,10}. These changes highlight the need for computational methods to accurately triage billable messages in accordance with policies, emphasizing clinical decision-making complexity rather than solely the time spent. This systematic review examines the computational use of SM data to enhance patient engagement, optimizing patient-provider communication, and reducing clinician burnout.

Recent advancements in computational tools and applications have shown significant potential to enhance the functionality of SM systems. NLP models have been employed to automate message classification, urgency triage, and the identification of medical decision complexity to improve clinical workflow efficiency^{21,23,28–30,33,39,41}. Additionally, applications have been developed to extract social determinants of health (SDOH) from messages, revealing prevalent unmet needs of patients such as financial strain, food insecurity, and housing instability^{33,42}. While BERT-based approaches have demonstrated high performance in multi-class classification, traditional ML models remain valuable in settings requiring interpretability and lower computational costs. Despite high model performance, challenges persist, particularly with class imbalances in SM datasets, which reduce algorithm sensitivity to underrepresented categories such as urgent messages and SDOH-related concerns. Moreover, the generalizability of existing tools is constrained by their development within a single healthcare institution, specialty, or patient demographic, limiting their applicability across diverse clinical settings. Since a single institution alone cannot generate sufficient evidence to fully address these limitations, future research should diversify datasets to improve model robustness and interoperability across varied patient populations⁴⁶. Transfer learning and federated learning approaches could enhance model adaptability across institutions while preserving data privacy⁴⁷. Furthermore, the integration of explainable AI methodologies has potential to enhance transparency and interpretability in clinical decision support, particularly for tasks with direct implications for patient care. Additionally, optimizing algorithmic trade-offs—such as maximizing recall for extracting social determinants and mental health indicators while maintaining high precision in triage systems—will be crucial for aligning computational advancements with practical clinical applications.

Beyond computational advancements, our literature screening revealed that some studies have developed theoretical frameworks based on SM corpus to support patient-provider communication. These frameworks complement computational methods in automated message processing, literacy profiling, and assessing patient engagement. One example is a theory-driven digital patient engagement model that established two robust sets of engagement measures, prioritizing factors that influence online patient interactions and using expert-reviewed data to validate engagement metrics⁴⁸. This framework provides a structured approach for evaluating how patients interact with SM platforms, and may enhance prediction performance of computational models to support automated triage and response prioritization. Another study addressed EHR message design by integrating cognitive and behavioral science theories to enhance patient portal SM comprehension⁴⁹. By evaluating different message formats, the study identified strategies to improve patient understanding, particularly among older adults, providing insights to inform computational developments on message simplification, readability prediction, and personalized messaging models. By integrating these frameworks with computational applications, future research could enhance clinical AI applications, ensuring that EHR systems design and implementation align with established theories of patient communication and engagement. As AI-based communication tools become increasingly embedded in clinical workflows, aligning application design with social, psychological, and behavioral theories is critical^{50,51}. These

frameworks provide structured guidance to bridge communication gaps, particularly in telehealth settings where engagement dynamics have evolved significantly over the past decades. Moreover, a major challenge in AI-driven messaging tools is their lack of interpretability, as many function as ‘black boxes’ with unclear decision-making processes. Theoretical frameworks contribute to greater transparency and ethical grounding by offering structured frameworks that make communication patterns more understandable and explainable. Therefore, integrating ethical guidelines, social-behavioral theories, and human-centered design principles is crucial to improving the explainability and trustworthiness of computational applications in patient–provider communication⁵².

Generative AI (GenAI) and Large Language Model (LLM) are gaining more attention in healthcare, particularly in the management of in-basket messages^{53,53–55}. Recent studies have evaluated the effectiveness of GenAI-generated responses compared to those written by physicians. Research indicates that while healthcare providers often prefer human-generated messages due to their higher readability and empathy levels^{56–58}, patients are generally satisfied with GenAI responses⁵⁹. This highlights a potential acceptance of GenAI in patient interactions, depending on the context and the expectations set for communication quality. Further investigation into GenAI has focused on its ability to handle inquiries of varying complexity⁶⁰. Advanced prompt engineering techniques and survey studies have been employed to gauge physicians' perceptions regarding the applicability and feasibility of using GenAI for message drafting^{61,62}. Additionally, pilot studies and quality improvement projects are actively exploring the practical implementation of GenAI in clinical settings^{55,63}. These preliminary findings highlight the need for ongoing advancements in GenAI capabilities, particularly in handling complex messages and enhancing readability, while prioritizing patient privacy to align with the principles of meaningful use^{1,2}. Our review of existing computational applications can provide valuable guidelines to inform the development of GenAI, ensuring alignment with clinical needs and patient-centered care principles. Addressing ethical considerations—including transparency in AI decision-making and safeguarding patient trust—will be critical for the responsible adoption of AI in patient–provider communication.

Conclusion

Advancements in the computational use of secure messaging (SM) data have contributed to the evolution of SM systems, enhancing message classification, urgency triage, and patient–provider communication. However, challenges remain in ensuring model adaptability, fairness, and seamless clinical integration. To refine AI-driven solutions, a deeper synergy between computational applications and communication theories is needed. In particular, as GenAI becomes more prevalent, it is essential to leverage its capabilities to handle complex messages and enhance readability while ensuring patient privacy, ethical integrity, and alignment with e-visit billing policies. Furthermore, rigorous assessment is required to evaluate how these technologies reshape clinical workflows, ensuring they support, rather than disrupt, patient–provider interactions. Moving forward, interdisciplinary collaboration will be key to bridging technical and theoretical gaps to develop policies that balance automation with clinical oversight. By aligning computational innovation with the complex nature of patient–provider communication, SM systems can better support efficient, equitable, and high-quality patient care.

References

1. Rights (OCR) O for C. HITECH Act Enforcement Interim Final Rule [Internet]. 2009 [cited 2024 Nov 5]. Available from: <https://www.hhs.gov/hipaa/for-professionals/special-topics/hitech-act-enforcement-interim-final-rule/index.html>
2. Blumenthal D, Tavenner M. The “Meaningful Use” Regulation for Electronic Health Records. *N Engl J Med*. 2010 Aug 5;363(6):501–4.
3. Eligible Professional’s Guide to STAGE 2 of the EHR Incentive Programs.
4. Say Goodbye to Paper and Pagers: Modernizing Clinical Communication | Epic [Internet]. [cited 2024 Nov 5]. Available from: <https://www.epic.com/epic/post/say-goodbye-to-paper-and-pagers-modernizing-clinical-communication-with-secure-chat/>
5. Secure Messaging Component [Internet]. [cited 2024 Nov 5]. Available from: <https://cernprphelp.cernerworks.com/Help/default.aspx?pg=11>
6. Ozkaynak M, Johnson S, Shimada S, Petrakis BA, Tulu B, Archambeault C, et al. Examining the Multi-level Fit between Work and Technology in a Secure Messaging Implementation. *AMIA Annu Symp Proc*.

- 2014;2014:954–62.
7. Haun JN, Alman AC, Melillo C, Standifer M, McMahon-Grenz J, Shin M, et al. Using Electronic Data Collection Platforms to Assess Complementary and Integrative Health Patient-Reported Outcomes: Feasibility Project. *JMIR Med Inf.* 2020;8(6):e15609.
8. Sisk BA, Bereitschaft C, Enloe M, Schulz G, Mack J, DuBois J. Oncology Clinicians' Perspectives on Online Patient Portal Use in Pediatric and Adolescent Cancer. *JCO Clin Cancer Inf.* 2023;7:e2300124.
9. Liu T, Zhu Z, Holmgren AJ, Ellimoottil C. National trends in billing patient portal messages as e-visit services in traditional Medicare. *Health Aff Sch.* 2024 Apr 4;2(4):qxae040.
10. Holmgren AJ, Grouse CK, Oates A, O'Brien J, Byron ME. Changes in Secure Messaging After Implementation of Billing E-Visits by Demographic Group. *JAMA Netw Open.* 2024;7(8):e2427053.
11. Holmgren AJ, Byron ME, Grouse CK, Adler-Milstein J. Association Between Billing Patient Portal Messages as e-Visits and Patient Messaging Volume. *JAMA.* 2023 Jan 24;329(4):339.
12. McGeady D, Kujala J, Ilvonen K. The impact of patient–physician web messaging on healthcare service provision. *Int J Med Inf.* 2008 Jan;77(1):17–23.
13. Wallwiener M, Wallwiener CW, Kansy JK, Seeger H, Rajab TK. Impact of electronic messaging on the patient–physician interaction. *J Telemed Telecare.* 2009 Jul;15(5):243–50.
14. Goldzweig CL, Towfigh AA, Paige NM, Orshansky G, Haggstrom DA, Beroes JM, et al. Systematic Review: Secure Messaging Between Providers and Patients, and Patients' Access to Their Own Medical Record: Evidence on Health Outcomes, Satisfaction, Efficiency and Attitudes [Internet]. Washington (DC): Department of Veterans Affairs (US); 2012 [cited 2023 Dec 29]. (VA Evidence-based Synthesis Program Reports). Available from: <http://www.ncbi.nlm.nih.gov/books/NBK100359/>
15. Kuo A, Dang S. Secure Messaging in Electronic Health Records and Its Impact on Diabetes Clinical Outcomes: A Systematic Review. *Telemed E-Health.* 2016 Sep;22(9):769–77.
16. Chen S, Guevara M, Moningi S, Hoebers F, Elhalawani H, Kann BH, et al. The impact of responding to patient messages with large language model assistance [Internet]. *arXiv*; 2023 [cited 2024 Sep 22]. Available from: <http://arxiv.org/abs/2310.17703>
17. Chen S, Guevara M, Moningi S, Hoebers F, Elhalawani H, Kann BH, et al. The effect of using a large language model to respond to patient messages. *Lancet Digit Health.* 2024 Jun 1;6(6):e379–81.
18. Liu S, McCoy AB, Wright AP, Carew B, Jenkins JZ, Huang SS, et al. Leveraging large language models for generating responses to patient messages—a subjective analysis. *J Am Med Inform Assoc.* 2024 May 20;31(6):1367–79.
19. Van Buchem MM, De Hond AAH, Fanconi C, Shah V, Schuessler M, Kant IMJ, et al. Applying natural language processing to patient messages to identify depression concerns in cancer patients. *J Am Med Inform Assoc.* 2024 Oct 1;31(10):2255–62.
20. Yang J, So J, Zhang H, Jones S, Connolly DM, Golding C, et al. Development and evaluation of an artificial intelligence-based workflow for the prioritization of patient portal messages. *JAMIA Open.* 2024;7(3):ooae078.
21. Si S, Wang R, Wosik J, Zhang H, Dov D, Wang G, et al. Students Need More Attention: BERT-based Attention Model for Small Data with Application to Automatic Patient Message Triage. In 2020. p. 436–56. Available from: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85146089595&partnerID=40&md5=a35243e9feabb73e7c34e6a8909bb02a>
22. Mermin-Bunnell K, Zhu Y, Hornback A, Damhorst G, Walker T, Robichaux C, et al. Use of Natural Language Processing of Patient-Initiated Electronic Health Record Messages to Identify Patients With COVID-19 Infection. *JAMA Netw Open.* 2023;6(7):e2322299.
23. Ren Y, Wu Y, Fan JW, Khurana A, Fu S, Wu D, et al. Automatic uncovering of patient primary concerns in portal messages using a fusion framework of pretrained language models. *J Am Med Inform Assoc.* 2024 Aug 1;31(8):1714–24.
24. Crossley SA, Balyan R, Liu J, Karter AJ, McNamara D, Schillinger D. Predicting the readability of physicians' secure messages to improve health communication using novel linguistic features: Findings from the ECLIPPSE study. *J Commun Heal.* 2020;13(4):1–13.
25. Schillinger D, Balyan R, Crossley SA, McNamara DS, Liu JY, Karter AJ. Employing computational linguistics techniques to identify limited patient health literacy: Findings from the ECLIPPSE study. *Health Serv Res.* 2021;56(1):132–44.
26. Song Q, Ni C, Warner JL, Chen Q, Song L, Rosenbloom ST, et al. Optimizing word embeddings for small datasets: a case study on patient portal messages from breast cancer patients. *Sci Rep.* 2024 Jul 12;14(1):16117.
27. De A, Huang M, Feng T, Yue X, Yao L. Analyzing Patient Secure Messages Using a Fast Health Care

- Interoperability Resources (FIHR)–Based Data Model: Development and Topic Modeling Study. *J Med Internet Res.* 2021 Jul 30;23(7):e26770.
28. Cronin RM, Fabbri D, Denny JC, Rosenbloom ST, Jackson GP. A comparison of rule-based and machine learning approaches for classifying patient portal messages. *Int J Med Inf.* 2017 Sep;105:110–20.
29. Mermin-Bunnell K, Zhu Y, Hornback A, Damhorst G, Walker T, Robichaux C, et al. Use of Natural Language Processing of Patient-Initiated Electronic Health Record Messages to Identify Patients With COVID-19 Infection. *JAMA Netw Open.* 2023 Jul 7;6(7):e2322299.
30. Sulieman L, Robinson JR, Jackson GP. Automating the Classification of Complexity of Medical Decision-Making in Patient-Provider Messaging in a Patient Portal. *J Surg Res.* 2020 Nov;255:224–32.
31. Chen J, Lalor J, Liu W, Druhl E, Granillo E, Vimalananda VG, et al. Detecting Hypoglycemia Incidents Reported in Patients’ Secure Messages: Using Cost-Sensitive Learning and Oversampling to Reduce Data Imbalance. *J Med Internet Res.* 2019 Mar 11;21(3):e11990.
32. Benda NC, Rogers C, Sharma M, Narain W, Diamond LC, Ancker J, et al. Identifying Nonpatient Authors of Patient Portal Secure Messages in Oncology: A Proof-of-Concept Demonstration of Natural Language Processing Methods. *JCO Clin Cancer Inform.* 2022 Dec;6(6):e2200071.
33. Cronin RM, Jackson GP. Automated Classification of Consumer Health Information Needs in Patient Portal Messages.
34. Ren Y, Wu D, Khurana A, Mastorakos G, Fu S, Zong N, et al. Classification of Patient Portal Messages with BERT-based Language Models. In: 2023 IEEE 11th International Conference on Healthcare Informatics (ICHI) [Internet]. Houston, TX, USA: IEEE; 2023 [cited 2025 Mar 3]. p. 176–82. Available from: <https://ieeexplore.ieee.org/document/10337114/>
35. A. P. Tafti, S. Fu, A. Khurana, G. M. Mastorakos, K. G. Poole, S. J. Traub, et al. Artificial intelligence to organize patient portal messages: a journey from an ensemble deep learning text classification to rule-based named entity recognition. In 2019. p. 1380–7.
36. Moon S, Wu Y, Doughty JB, Wieland ML, Philpot LM, Fan JW, et al. Automated Identification of Patients’ Unmet Social Needs in Clinical Text Using Natural Language Processing. *Mayo Clin Proc Digit Health.* 2024;2(3):411–20.
37. Balyan R, Crossley SA, Brown W, Karter AJ, McNamara DS, Liu JY, et al. Using natural language processing and machine learning to classify health literacy from secure messages: The ECLIPPSE study. Grabar N, editor. *PLOS ONE.* 2019 Feb 22;14(2):e0212488.
38. Crossley SA, Balyan R, Liu J, Karter AJ, McNamara D, Schillinger D. Developing and Testing Automatic Models of Patient Communicative Health Literacy Using Linguistic Features: Findings from the ECLIPPSE study. *Health Commun.* 2021;36(8):1018–28.
39. De A, Huang M, Feng T, Yue X, Yao L. Analyzing Patient Secure Messages Using a Fast Health Care Interoperability Resources (FIHR)-Based Data Model: Development and Topic Modeling Study. *J Med Internet Res.* 2021;23(7):e26770.
40. Sulieman L, Gilmore D, French C, Cronin RM, Jackson GP, Russell M, et al. Classifying patient portal messages using Convolutional Neural Networks. *J Biomed Inf.* 2017;74:59–70.
41. Yang J, So J, Zhang H, Jones S, Connolly DM, Golding C, et al. Development and evaluation of an artificial intelligence-based workflow for the prioritization of patient portal messages. *JAMIA Open.* 2024 Jul 1;7(3):o0ae078.
42. Moon S, Wu Y, Doughty JB, Wieland ML, Philpot LM, Fan JW, et al. Automated Identification of Patients’ Unmet Social Needs in Clinical Text Using Natural Language Processing. *Mayo Clin Proc Digit Health.* 2024 Sep;2(3):411–20.
43. Padman R, Shevchik G, Paone S, Dolezal C, Cervenak J. eVisit: a pilot study of a new kind of healthcare delivery. *Stud Health Technol Inf.* 2010;160(Pt 1):262–6.
44. Parpia C, Moore C, Beatty M, Miranda S, Adams S, Stinson J, et al. Evaluation of a Secure Messaging System in the Care of Children With Medical Complexity: Mixed Methods Study. *JMIR Form Res.* 2023;7:e42881.
45. Huang M, Fan J, Prigge J, Shah ND, Costello BA, Yao L. Characterizing Patient-Clinician Communication in Secure Medical Messages: Retrospective Study. *J Med Internet Res.* 2022;24(1):e17273.
46. Betancourt JR, Maina AW. The Institute of Medicine report “Unequal Treatment”: implications for academic health centers. *Mt Sinai J Med N Y.* 2004 Oct;71(5):314–21.
47. Teo ZL, Jin L, Liu N, Li S, Miao D, Zhang X, et al. Federated machine learning in healthcare: A systematic review on clinical applications and technical architecture. *Cell Rep Med.* 2024 Feb;5(2):101419.
48. Garvin LA, Simon SR. Prioritizing Measures of Digital Patient Engagement: A Delphi Expert Panel Study. *J Med Internet Res.* 2017 May 26;19(5):e182.

49. Morrow D, Hasegawa-Johnson M, Huang T, Schuh W, Azevedo RFL, Gu K, et al. A multidisciplinary approach to designing and evaluating Electronic Medical Record portal messages that support patient self-care. *J Biomed Inf.* 2017;69:63–74.
50. Uribe-Toril J, Ruiz-Real JL, Nievas-Soriano BJ. A Study of eHealth from the Perspective of Social Sciences. *Healthcare.* 2021 Jan 21;9(2):108.
51. Pagliari C. Design and evaluation in eHealth: challenges and implications for an interdisciplinary field. *J Med Internet Res.* 2007 May 27;9(2):e15.
52. Van Velsen L, Ludden G, Grünloh C. The Limitations of User-and Human-Centered Design in an eHealth Context and How to Move Beyond Them. *J Med Internet Res.* 2022 Oct 5;24(10):e37341.
53. Rodriguez DV, Andreadis K, Chen J, Gonzalez J, Mann D. Development of a GenAI-Powered Hypertension Management Assistant: Early Development Phases and Architectural Design. In 2024. p. 350–9. Available from: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85203709593&doi=10.1109%2fchi61247.2024.00052&partnerID=40&md5=dc9f4e77f43f77a342ff9a0e6232284>
54. Clusmann J, Kolbinger FR, Muti HS, Carrero ZI, Eckardt JN, Laleh NG, et al. The future landscape of large language models in medicine. *Commun Med.* 2023 Oct 10;3(1):141.
55. Garcia P, Ma SP, Shah S, Smith M, Jeong Y, Devon-Sand A, et al. Artificial Intelligence–Generated Draft Replies to Patient Inbox Messages. *JAMA Netw Open.* 2024 Mar 20;7(3):e243201.
56. Reynolds K, Nadelman D, Durgin J, Ansah-Addo S, Cole D, Fayne R, et al. Comparing the quality of ChatGPT- and physician-generated responses to patients’ dermatology questions in the electronic medical record. *Clin Exp Dermatol.* 2024 Jul 1;49(7):715–8.
57. Baxter SL, Longhurst CA, Millen M, Sitapati AM, Tai-Seale M. Generative artificial intelligence responses to patient messages in the electronic health record: early lessons learned. *JAMIA Open.* 2024 Jul 1;7(2):ooae028.
58. Small WR, Wiesenfeld B, Brandfield-Harvey B, Jonassen Z, Mandal S, Stevens ER, et al. Large Language Model–Based Responses to Patients’ In-Basket Messages. *JAMA Netw Open.* 2024 Jul 16;7(7):e2422399.
59. Kim J, Chen ML, Rezaei SJ, Liang AS, Seav SM, Onyeka S, et al. Patient Perspectives on Large Language Model Responses to Patient Messages [Internet]. 2024 [cited 2024 Sep 22]. Available from: <https://www.ssrn.com/abstract=4867523>
60. Scott M, Muncey W, Seranio N, Belladelli F, Del Giudice F, Li S, et al. Assessing Artificial Intelligence–Generated Responses to Urology Patient In-Basket Messages. *Urol Pract.* 2024 Sep;11(5):793–8.
61. Kaur A, Budko A, Liu K, Eaton E, Steitz B, Johnson KB. Automating Responses to Patient Portal Messages Using Generative AI [Internet]. *medRxiv*; 2024 [cited 2024 Jun 19]. p. 2024.04.25.24306183. Available from: <https://www.medrxiv.org/content/10.1101/2024.04.25.24306183v1>
62. Yan S, Knapp W, Leong A, Kadkhodazadeh S, Das S, Jones VG, et al. Prompt engineering on leveraging large language models in generating response to InBasket messages. *J Am Med Inform Assoc.* 2024 Jul 19;31(10):2263–70.
63. Tai-Seale M, Baxter SL, Vaida F, Walker A, Sitapati AM, Osborne C, et al. AI-Generated Draft Replies Integrated Into Health Records and Physicians’ Electronic Communication. *JAMA Netw Open.* 2024 Apr 15;7(4):e246565.