Eliminating Risk Involved in Using ChatGPT for Clinical Decision Support System

Kenisha D'Costa(✉) and Vijay Kanabar

Boston University, Metropolitan College, 1010 Commonwealth Ave., Boston, MA, USA
{dcostk2,kanabar}@bu.edu

Abstract. The paper discusses the integration of Generative AI, specifically ChatGPT, in Clinical Decision Support Systems (CDSS). These intelligent models have applications that are being widely adopted in every industry, even in a clinical setting. It helps optimize Clinical Decision Support (CDS) and provides unique suggestions. It can recognize possible modifications to alert logic and even help clinicians recommend CDS development. However, AI use does have certain risks, such as hallucination. ChatGPT, a very well-known generative AI model, can handle large amounts of data, and it also makes it convenient for clinicians to review diagnostic suggestions during the process of CDS. This paper examines the possible challenges of implementing ChatGPT in a clinical setting. Our research introduces a new model, essentially an AI loop called Spiral Human-In-The-Generative (Spiral HITL), which seeks to mitigate risks involved with AI applications in CDS.

Keywords: Generative Artificial Intelligence (AI) · Clinical Decision Support (CDSS) · Spiral HITL model · ChatGPT

1 Introduction

OpenAI created ChatGPT, a large language model meant to converse with people conversationally. It is a large language model with the Generative Pre-trained Transformer (GPT) architecture as its foundation. Given understandable human input, it can comprehend and produce writing that humans can interpret. With ChatGPT [1], users may engage in discussions, ask questions, get help, or have fun chatting. It can generate text that is grammatically correct and coherent. This drew public attention to its potential applications across various industries, including education and healthcare. Generative AI tools possess diverse capabilities, from enhancing topic comprehension to aiding in project leadership, management, and control [1–3].

ChatGPT generates natural language content using deep learning techniques based on the GPT architecture. Supervised and reinforcement learning techniques have improved the model [3, 4]. When a prompt is entered, it gathers information from all available sources, feeds it into a transformer model, maps the connections between various data points, and makes informed recommendations about text patterns [1].

T. Zlateva and G. Tuparov (Eds.): CSECS 2024, LNICST 609, pp. 201–215, 2025.
https://doi.org/10.1007/978-3-031-84312-9_14

AI models successfully demonstrate continuous improvement through supervised and reinforcement learning. This has garnered significant public attention, highlighting its potential applications across various industries.

Note while there a wide range of large language models we focus specifically on ChatGPT due to its prominent and early introduction into the marketplace. Furthermore, it is widely recognized for its capability to effectively interface with human input to enhance Clinical Decision Support (CDS). With this effective AI tool, doctors may learn new techniques for enhancing clinical decision-making and receive guidance and support. Large volumes of medical data can be analyzed, patterns can be found, and best clinical practices may be gained through its application in research and development. Hospitals should consider integrating ChatGPT into their practices for its benefits. Additionally, employing ChatGPT should be viewed as an aid rather than a substitute for clinicians' clinical practices [4, 5].

ChatGPT can automate data gathering for study and lessen the drawbacks of manual extraction, time consumption, and human mistakes. For example, to assist patients in better comprehending their therapy, a researcher may instruct the system to create a discharge statement for them once their treatment is complete. It might also be used to immediately respond to questions from clinicians by assessing the patient data that is already accessible and offering a thorough response. This might encourage more effective patient treatment, especially when a doctor needs a summary of the patient's therapeutic journey [2].

In summary, foundational models can automate data gathering and generate patient discharge statements or summaries. They can significantly enhance the efficiency and accuracy of patient care by reducing manual extraction errors and providing timely, comprehensive responses to clinicians' inquiries.

2 Background and Motivation

2.1 Importance of Clinical Decision Support

Clinical Decision Support CDS provides valuable information about the patient under study that a clinician can use to understand the treatment process. It connects the clinician to the knowledge base with all the stored clinical information [6] (Fig. 1).

Clinical Decision Support System (CDSS) aims to enhance healthcare delivery by enriching medical decision-making with targeted clinical knowledge, patient data, and other health information. These days, clinicians mostly employ CDSSs at the point of care when they integrate their expertise with the data or recommendations from the CDSS. However, more and more CDSSs are being created that can use observations and data that would otherwise be unavailable or difficult for people to understand [7].

Implementing sound clinical decision support systems enhances medical professionals' decision-making, improves care standards, and facilitates better disease management and prevention, fostering increased patient participation through interactive, collaborative tools.

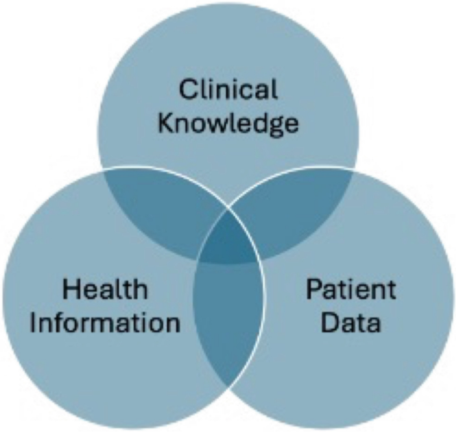

Fig. 1. Clinical Decision Support Components

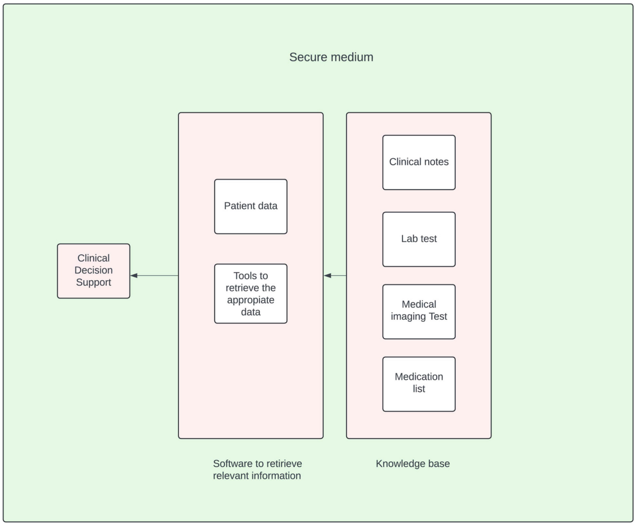

Fig. 2. Overview of Clinical Decision Support System

Figure 2 shows a high-level overview of clinical decision support software. A clinician can use several sources, such as lab tests, imaging tests, and other sources of information, to aid them in making an appropriate diagnosis. The patient's information can be obtained using the tools embedded in the Electronic Health Record (EHR).

CDSS can facilitate increasing patient participation in clinical decision-making by providing interactive tools that let patients go through relevant information that can promote collaborative decision-making [8]. There are several advantages to implementing sound clinical decision support. They serve as vital resources that assist medical professionals in making knowledgeable decisions, raising the benchmark for overall care, and facilitating improved illness management and prevention [6, 9, 10].

A Clinical Decision Support System supports:

- Better use of evidence – It provides personalized clinical expertise for patients and intelligent and careful application of the most recent best evidence from clinical care research. This is known as evidence-based medicine. This method calls for clinical judgment and allows for missing, insufficient, or poor-quality evidence. CDS guarantees they have access to the latest studies and expert suggestions and helps them create approaches based on the most recent findings [6, 9, 10].
- Improved accuracy - CDS boosts efficiency by giving physicians and decision-makers access to real-time, data-driven insights that facilitate more precise, timely, and informed decision-making. The Laboratory Information Management System (LIMS) may be integrated with it to manage related information efficiently. This can manage samples and associated data, integrate instruments, and automate lab procedures [6, 9, 10].
- Personalized care plans - Large-scale data sets and sophisticated algorithms are used with CDSSs to provide more precise and customized suggestions. IBM Watson Health and Google's DeepMind are two examples of how AI and ML have the potential to revolutionize healthcare decision-making [11].

Kabachinski [12] summarizes that the healthcare industry has three major simultaneous developments in knowledge management and information technology (IT), "The first is data collection, i.e., the adoption and "meaningful use" of electronic health records (EHRs). Second is data sharing, as exemplified by health information exchanges (HIEs). Finally, there is data analysis as seen in enterprise data warehouse (EDW) and clinical decision support system (CDSS) tools."

2.2 Issues with CDSS

As we have just seen, clinical decision support systems (CDSS) improve patient care, reduce errors, and make things more efficient for doctors. These systems give doctors evidence-based advice when they see a patient. However, there are challenges to getting these systems adopted and working well, such as privacy concerns, making them work with other computer systems, and getting doctors to use them. Even though CDSS has much potential, it's still difficult to get them widely used and to work effectively [11].

Existing AI CDSS systems need more structured integration of human expertise during development, leading to potential inaccuracies and usability issues. Despite the potential of advanced clinical decision support systems (CDSS) to transform clinical practice through sophisticated algorithms and patient-specific insights, challenges such as inconsistent error detection by EHR systems, clinician fatigue due to increased workload and alert fatigue, and the complexity of data integration and relevancy of alerts must be addressed to enhance their effectiveness and user acceptability.

Medication alert errors occur when Clinical Decision Support Systems (CDS) generate incorrect alerts about medication use, potentially leading to inappropriate medication use or patient harm. Common types of errors include [13]:

- False positives lead to alert fatigue, where healthcare providers ignore essential alerts.
- False negatives trigger an alert, possibly resulting in harmful medication errors.
- Incorrect alerts pertain to incorrect drug interactions or dosages.
- Overriding alerts without proper evaluation leads to medication errors.

Research published in JAMA Network Open found that CDS warnings miss up to 33% of therapeutic mistakes. Researchers from Brigham and Women's Hospital, Harvard University, and University of Utah Health discovered that while Electronic Health Record (EHR) systems are inconsistent in identifying mistakes, patients are nevertheless placed in dangerous and sometimes fatal circumstances [14] (Fig. 3).

Fig. 3. Challenges Faced with Existing CDSS

Clinician fatigue - Clinicians have experienced several effects from CDS, such as changing workflow, increased workload or additional tasks, alert/alarm fatigue, work spilling over into personal time, diminished autonomy, and worry about the medico-legal repercussions of CDS recommendations [15].

Increased complexity - With healthcare moving forward, decision support system development is taking center stage. Advanced CDSS has the potential to completely transform clinical practice because of its advanced algorithms and patient-specific characteristics. Through integrating laboratory data, mitigating medication interactions, and navigating sophisticated contraindications, these systems provide medical professionals with deep insights that are critical for making well-informed decisions. Nevertheless, their effectiveness depends on overcoming obstacles related to data integration, alert relevancy, and user acceptability, highlighting the continuous pursuit of improvement in decision-support technology [16].

Figure 4 shows the complexity of the clinical decision support system. They are stored in different formats, making it complicated to retrieve at any given point in time.

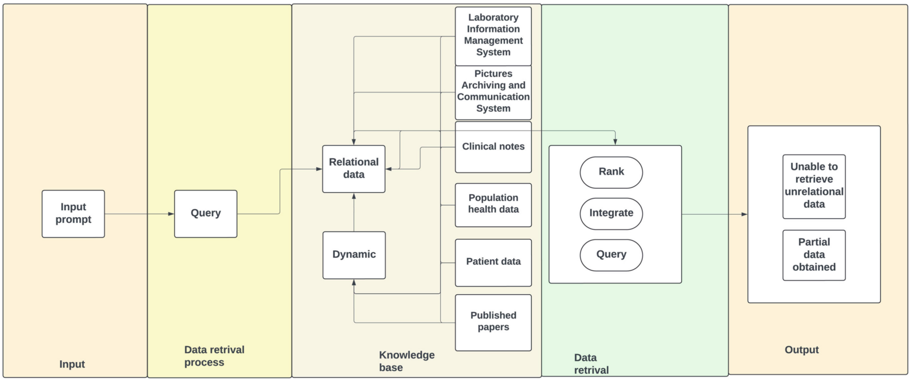

Fig. 4. Complexity in Clinical Decision Support

As the complexity of the data increases, another drawback observed is that certain critical information gets suppressed; it also becomes more accessible for a clinician to brainstorm diagnoses if the necessary information is unavailable.

2.3 Integrating ChatGPT with Clinical Decision Support

CDS mainly uses static rules and algorithms, which might not consider each patient's particular and complicated set of attributes; by offering recommendations based on patient data, clinical guidelines, and best practices in real-time, generative AI can assist medical professionals in making better judgments. This can boost patient outcomes, lower medical mistakes, and increase diagnosis accuracy [17].

According to Mass General Brigham researchers, ChatGPT made clinical decisions with an accuracy of 71.7% across all 36 clinical scenarios. Possible diagnoses were generated by ChatGPT, which then made judgments on care management and final diagnosis [18].

When determining a final diagnosis, ChatGPT performed the best in research partially supported by the National Institute of General Medical Sciences; the model achieved 77% accuracy.

Figure 5 demonstrates how ChatGPT can be trained on information for individual patients and assist in making suggestions.

In summary, integrating ChatGPT with Clinical Decision Support can enhance decision-making by providing real-time recommendations based on patient data, clinical guidelines, and best practices, potentially improving patient outcomes, reducing medical errors, and increasing diagnosis accuracy, with research showing ChatGPT achieving an overall accuracy of 71.7% and up to 77% in determining final diagnoses in clinical scenarios.

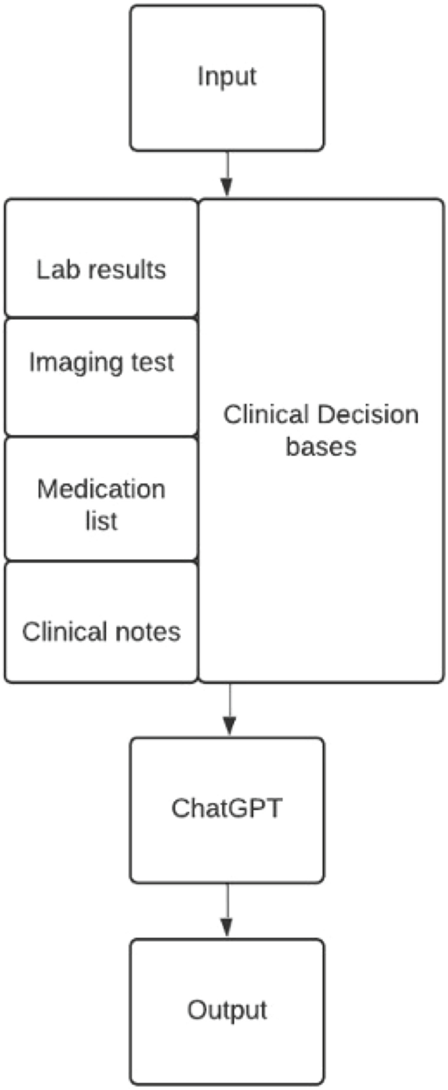

Fig. 5. Integrating ChatGPT in Clinical Decision Making

3 Literature Review

3.1 Background

A study by Liu et al. [19] at Vanderbilt University Medical Center aimed to evaluate ChatGPT's suggestions for medical alerts, comparing them with expert recommendations by categorizing and rating each suggestion; using GPT-3.5, the suggestions were based on alerts from the Clickbusters program, reviewed by experienced clinical informaticians who were blinded to the source of the recommendations.

In the study [19], the suggestions of ChatGPT and experts in the field of several medical alerts were examined, then categorized based on several indicators, and ratings were given for each suggestion. The suggestions were based on the alerts by the Clickbusters program, an initiative by Vanderbilt Clinical Information Center (VCLIC) that focuses on improving safety and quality concerns using Epic. The suggestions were generated by recruited experts from VUMC UT Southwestern Medical Center and ChatGPT (GPT 3.5), and they were grouped by the types of alerts that needed improvement in the Clickbusters program. These were later reviewed by five clinical informaticians with experience using CDS and EHRs for 12–15 years. These informaticians were unaware whether ChatGPT or humans had made the recommendations. The recommendations that were given for evaluation were a blend of suggestions from AI and humans.

3.2 Method and Findings

The research evaluated each recommendation using a 5-point Likert scale, with categories like Comprehension, Relevance, Usefulness, Acceptance, Workflow, Redundancy, Inversion, and Prejudice. Clinicians' feedback was collected via REDCap surveys, which included each alert's description, reasoning, and additional information. REDCap (Research Electronic Data Capture) is a secure web application for data collection and management in research studies. It is widely used for creating and managing online surveys and databases.

The stakeholders in the study are clinicians from VUMC and UT Southwestern Medical Center and experts from departments like Internal Medicine, Anesthesiology, Pharmacy, Pediatrics, and Emergency Medicine.

Among the thirty-six AI-generated proposals, 75% received an overall score of three or above, with a minimum of 2.6 and a maximum of 4.15. 3.3 ± 0.5 was the average score. Meanwhile, the suggestions from the clinicians received a higher score than those from the generative tool. Table 2 compares AI-generated and human-generated suggestions based on several categories [19].

AI-generated suggestions could complement human efforts in optimizing CDS alerts, identifying potential improvements, and supporting implementation. ChatGPT shows promise in using large language models and reinforcement learning to enhance CDS alert logic and other areas involving complex clinical logic. ChatGPT's scores matched those of clinicians, offering unique perspectives. With further training, AI can be effectively used in CDS.

3.3 Method and Findings

The research evaluated each recommendation using a 5-point Likert scale, with categories like Comprehension, Relevance, Usefulness, Acceptance, Workflow, Redundancy, Inversion, and Prejudice. Clinicians' feedback was collected via REDCap surveys, which included each alert's description, reasoning, and additional information. REDCap (Research Electronic Data Capture) is a secure web application for data collection and management in research studies. It is widely used for creating and managing online surveys and databases.

Even though ChatGPT had a high level of understanding and relevance, the researchers Liu et al. noticed certain risks, such as:

- The model needed to be made aware of the complete EHR operation, generating a few suggestions not meant for the current process.
- Hallucination
- The information generated was partially correct.
- Suggestions deviating from expert opinion.

The implications are that the AI model can generate solutions that are not applicable or suitable for the specific processes within EHR operations and provide partially correct solutions or outright incorrect results. These observations highlight the necessity for continuous improvement and rigorous validation of AI models in specialized domains.

4 Spiral Human-In-The-Loop Model

From the research presented so far, it is evident that alerts' relevance, accuracy, and usefulness can be improved if we combine the suggestions obtained by ChatGPT and clinicians. Barry Boehm's spiral model, a popular model in software development, specifically for agile software development, inspires us to refine the current AI DSS approach to make it more reliable [20]. In Boehm's spiral model, the software process is executed as a series of iterating steps going from one phase to another; it is depicted as a spiral. Every spiral loop represents a different stage of the software development process.

Next, let us consider the Human-In-The-Loop (HITL) approach; here, humans participate in both the algorithm's training and testing phases, and this model combines supervised machine learning and expert recommendation techniques [21].

Integrating the Spiral model with the HITL approach in an AI loop mitigates risks, eliminates Gen AI's shortcomings, and collectively improves CDS. This section describes the combination of the Spiral model and HITL approach. Our model is called *Spiral HITL* for convenience. The proposed Spiral HITL model demonstrates the following benefits:

- Enhanced Accuracy: By incorporating human expertise at each stage, the model can ensure that AI-generated recommendations are validated and accurate, reducing diagnostic errors.
- Iterative Improvement: The spiral model's iterative nature allows continuous refinement and adaptation.
- Risk Mitigation: Continuous risk analysis and assessment help identify and address potential issues early, improving the safety and reliability of the CDS.
- Personalized Care: Human input in the loop ensures that AI recommendations are tailored to individual patient needs, leading to better patient outcomes and more customized treatment plans.
- Regulatory Compliance and Acceptance: The structured development phases ensure the model meets regulatory standards, from prototyping to FDA approval. Combining AI with human oversight can enhance trust among healthcare providers, making them more likely to adopt and rely on the system (Fig. 6).

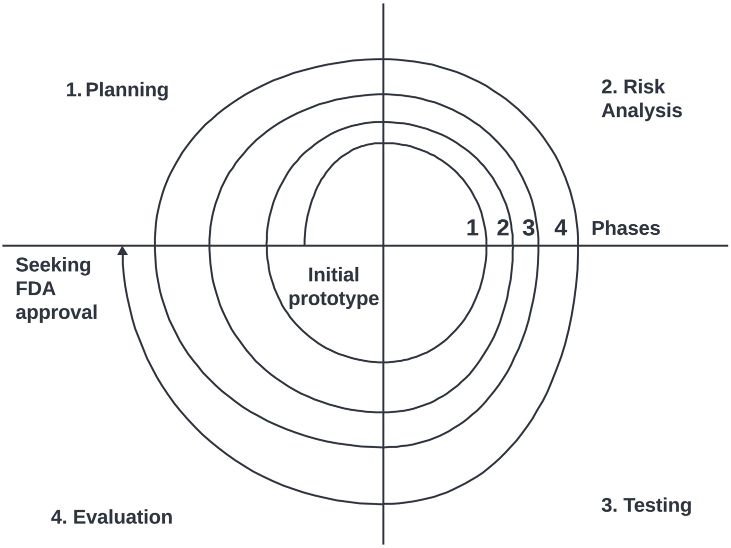

Fig. 6. Spiral Humans-In-The-Loop (Spiral HITL) Model

The detailed Spiral HITL model is presented in Table 2. The development takes four phases, each going through stages: planning, risk analysis, testing, and evaluation (Table 1).

Table 1. Ranking of Suggestions Generated by AI and Humans Based on Seven Categories

Category	Scores of AI Generated Suggestions	Scores of Human Generated Suggestions
Understanding	Agree	Agree
Relevance	Agree	Agree
Acceptance	Neither agree nor disagree	Agree
Workflow	Strongly disagree	Neither agree nor disagree
Bias	Strongly disagree	Strongly disagree
Inversion	Disagree	Strongly disagree
Redundancy	Disagree	Strongly disagree

Table 2. A High-Level Overview of the Various Phases in the Development Lifecycle

Stages	Phase 1	Phase 2	Phase 3	Phase 4
Planning	Business Analyst: Understanding the needs of clinicians	Designing the model (version 1), keeping in mind the risk and retrospective analysis from phase	Designing the model (version 2), based on the evaluation from phase 2	Clinical trials of the human in the gen AI loop model
Risk Analysis	High-level analysis of the compliance and uncertainties involved	Conducting Failure Mode Effect Analysis and coming up with Good Machine Learning Practices	Having a well-defined risk register	Validating the risk in the risk register
Testing	Developing a simulation candidate to test out Humans in the AI loop	Testing the model with sample data and expert suggestions	Testing out the human in the generative	Conducting clinical trials
Evaluation	Eliciting the resources needed for the project and retrospection of the simulation model	Retrospection of the model, gathering data for Cost of Quality (CoQ) and First Time Right (FTR)	Retrospection and preparing for clinical trials	Seeking FDA approval

The processes occur in spiral increments, progressing through several development stages from the prototype phase to seeking FDA approval . Each phase is detailed below:

Phase 1

Planning:

This initial stage involves business analysis to elicit requirements based on clinicians' needs and begin the product planning process, including resource allocation and budgeting for the entire development.

Risk Analysis:

All potential risks, including security concerns, hallucinations, and incorrect information generation, are identified at this stage. Security issues are delegated to the security team, while measures are implemented to prevent hallucinations and misinformation. Suggestions generated by the model will be reviewed by experts, and a prompt template (see Fig. 7) will be used to ensure accurate information delivery.

Testing:
The prototype undergoes both white box and black box testing. The software development team devises test cases, initially validated through white box testing by an engineer. Subsequently, black box testing is performed by another engineer unfamiliar with the model. Reports are documented and stored in the product lifecycle management database.

Evaluation:
This final stage of Phase 1 involves a recap of the entire phase, documenting best practices and identifying areas for improvement. The team is credited for successfully completing Phase 1.

The illustrated prompt in Fig. 7 asks ChatGPT to play the clinician role.

According to AAP guidelines, "Clinicians should not administer albuterol to infants and children diagnosed with bronchiolitis."

Here are the current inclusion requirements for treatment:

- Inpatient bronchiolitis in the past 14 days, with an order set utilized, or if bronchiolitis was added to the issue list.
- Age inclusion is "up to 2 years"; either chest x-rays or albuterol are required.

As a clinician, please analyze this patient's treatment and provide suggestions to support your reasoning. Additionally, are there any other exclusions that should be considered?

Fig. 7. Sample Prompt Template

Phase 2

Planning:
Product design is refined based on lessons learned from the Phase 1 prototype. Subsequent processes are briefly discussed, and any necessary adaptations or foreseeable risks are addressed.

Risk Analysis:
Failure Mode Effect Analysis (FMEA) and Good Machine Learning Practices are documented and stored in the Product Lifecycle Management (PLM) database. FMEA identifies and manages potential issues to prevent adverse outcomes. Good Machine Learning Practices ensure adherence to industry standards when using Generative AI in a hospital setting.

Testing:
The testing process from Phase 1 is repeated. Reports and documentation are reviewed by the team lead, software developer, and quality assurance lead before being stored in the PLM.

Evaluation:
Like Phase 1, a retrospective analysis is conducted. Cost of Quality and First Time Right information is documented and reviewed by stakeholders to improve future processes.

Cost of Quality assesses resource allocation to minimize quality issues, while First Time Right ensures tasks are completed accurately on the first attempt, reducing costs and improving quality.

Phase 3
Planning:
This crucial phase prepares the model for clinical trials and FDA approval. Design is based on Phases 1 and 2 evaluations, ensuring all risks are documented and mitigated.

Risk Analysis:
Future updates to the model are recorded and stored in the PLM. Design History Files documenting the product's development and history are also maintained.

Testing:
Consistent with earlier phases, the model undergoes testing, with results documented.

Evaluation:
A retrospective analysis is conducted, and necessary updates are documented. Additional phases may be proposed to implement these changes.

Phase 4
Planning:
Clinical trials are conducted, testing the model for specific clinical decision support.

Risk Analysis:
Documented risks are validated.

Testing:
Planned clinical trials are carried out, reported, and documented.

Evaluation:
Upon completion of this phase, the model seeks FDA approval. Once accepted, it will be implemented in hospitals.

5 Spiral HITL Closing Remarks

This model makes it easy to identify the risk involved and makes it convenient to incorporate changes needed while working on humans in the generative AI loop model. For example, if the need to change the generative AI model arises, it can be done by planning it for the appropriate stage.

It involves proper risk management and takes feedback iteratively. Thus ensuring everyone is engaged across every stage of its development. A generative AI loop model increases transparency with the AI-made decisions and ensures reliability and accuracy. Generative AI and the experts can work together to improve the accuracy of diagnosis and eliminate risk.

To err is human – even though this model emphasizes a suggestion from humans and AI combined, humans may tend to depend on AI's decisions in real-world applications. While this model integrates suggestions from both humans and AI, there is a risk that humans may overly rely on AI decisions in real-world applications. The model's effectiveness heavily depends on thorough risk assessment and evaluation stages; if these are not properly conducted, the project's outcomes could be negatively impacted. Therefore, it is crucial to rigorously evaluate risks at the project's outset to minimize workload and potential issues later.

Challenges and opportunities for future research are listed below:

- Ensuring effective and efficient collaboration between AI developers and healthcare professionals.
- Addressing potential biases in AI algorithms to ensure fair and equitable healthcare delivery.
- Maintaining data privacy and security throughout the iterative development process.

The Retrieval-Augmented Generation (RAG) approach can be a valuable addition to further mitigate AI decision risks in the Spiral HITL model. RAG combines the strengths of retrieval-based and generative models, leveraging large-scale external knowledge bases to enhance the quality and accuracy of AI-generated responses.

6 Conclusion

AI is a powerful tool, and it has several applications. If we train the model correctly, it can handle large volumes of EHR data. This paper introduces the spiral HITL model, which mitigates the risk of AI DSS systems. The Spiral HITL model simplifies risk identification and facilitates necessary changes while working with humans in the generative AI loop. For instance, any need to modify the generative AI model can be strategically planned for the appropriate development stage.

The model incorporates robust risk management and iterative feedback, ensuring continuous engagement and improvement at every development stage. This iterative process enhances transparency and reliability in AI decisions, as generative AI and human experts collaborate to improve diagnostic accuracy and mitigate risks effectively.

Despite emphasizing enhancing human and AI integration, there is a concern that we may become overly dependent on AI decisions. The model's effectiveness hinges on thorough risk assessment and evaluation stages; adequate execution of these stages can positively impact the outcomes.

References

1. ChatGPT model (2024). https://www.openai.com/
2. Janssen, B.V., Kazemier, G., Besselink, M.G.: The use of ChatGPT and other large language models in surgical science. BJS Open **7**(2) (2023). https://doi.org/10.1093/bjsopen/zrad032
3. Kanabar, V., Wong, J.: AI Revolution in Project Management. Pearson (2024)
4. Liu, J., Wang, C.L.S.: Utility of ChatGPT in clinical practice. J. Med. Internet Res. (2023). https://doi.org/10.2196/48568
5. Bhattacharya, K., Bhattacharya, A.S., Bhattacharya, N., Yagnik, V.D., Garg, P., Kumar, S.: ChatGPT in surgical practice—a new kid on the block. Indian J. Surg. **85**, 1346–1349 (2023). https://doi.org/10.1007/s12262-023-03727-x

6. Beeler, P.E., Bates, D.W., Hug, B.L.: Clinical decision support systems. Swiss Med. Wkly. **144**(5152), w14073 (2014). https://doi.org/10.4414/smw.2014.14073
7. Sutton, R.T., Pincock, D., Baumgart, D.C., Sadowski, D.C., Fedorak, R.N., Kroeker, K.I.: An overview of clinical decision support systems: benefits, risks, and strategies for success. npj Digit. Med. **3**(1), 17, 2020/02/06 (2020). https://doi.org/10.1038/s41746-020-0221-y
8. Sim, I., et al.: Clinical decision support systems for the practice of evidence-based medicine (in eng). J. Am. Med. Inform. Assoc. **8**(6), 527–34 (2001). https://doi.org/10.1136/jamia.2001.0080527
9. Folio, N.: Benefits of Clinical Decision Support System. https://digitalhealth.folio3.com/blog/benefits-of-clinical-decision-support-system/. Accessed 4 Jul 2024
10. Berner, E.S., (ed.) Clinical Decision Support Systems. Springer (2016)
11. Chen, Z., et al.: Harnessing the power of clinical decision support systems: challenges and opportunities. Open Heart **10**(2), e002432 (2023). https://doi.org/10.1136/openhrt-2023-002432
12. Kabachinski, J.: A look at clinical decision support systems. Biomed. Instrum. Technol. (2013). https://doi.org/10.2345/0899-8205-47.5.432
13. Mondul, A., Kong, M.: Medication error. In: Agrawal, A., Bhatt, J. (eds.) Patient Safety: A Case-based Innovative Playbook for Safer Care, pp. 151–164. Springer International Publishing, Cham (2023)
14. Jason, C.: CDS Alert Fatigue Runs Rampant in Ambulatory Clinics. TechTarget.com. https://www.techtarget.com/searchhealthit/news/366579210/CDS-Alert-Fatigue-Runs-Rampant-in-Ambulatory-Clinics. Accessed 10 Feb 2024
15. Jankovic, I., Chen, J.H.: Clinical decision support and implications for the clinician burnout crisis (in eng). Yearb. Med. Inform. **29**(1), 145–154 (2020). https://doi.org/10.1055/s-0040-1701986
16. Marco. D.: Complexity of clinical decision support systems. https://pharmafeatures.com/complexity-of-clinical-decision-support-systems/. Accessed 12 Feb 2024
17. Campos, B.: Navigating the future: The impact of generative AI and large language models in healthcare. HIMSS Iberian Community. https://www.himss.org/sites/hde/files/media/file/2024/04/05/himss_article_iberiancommunity.pdf. Accessed 3 July 2024
18. Fox, A.: ChatGPT scored 72% in clinical decision accuracy, MGB study shows. HIMSS Media Publication. https://www.healthcareitnews.com/news/chatgpt-scored-72-clinical-decision-accuracy-mgb-study-shows. Accessed 29 Aug 2023
19. Liu, S., et al.: Using AI-generated suggestions from ChatGPT to optimize clinical decision support. J. Am. Med. Inform. Assoc. **30**(7), 1237–1245 (2023). https://doi.org/10.1093/jamia/ocad072
20. Boehm, B.W.: A spiral model of software development and enhancement. Computer **21**(5), 61–72 (1988). https://doi.org/10.1109/2.59
21. Telus. Human-in-the-loop. https://www.telusinternational.com/glossary/human-in-the-loop. Accessed 4 July 2024
22. QAPI. Guidance for Performing Failure Mode and Effects Analysis with Performance Improvement Projects. https://www.cms.gov/medicare/provider-enrollment-and-certification/qapi/downloads/guidanceforfmea.pdf. Accessed 22 Mar 2024
23. ASQ. Cost of Quality. https://asq.org/quality-resources/cost-of-quality. Accessed 4 July 2024
24. Operations1. What does first time right mean?. https://operations1.com/en/glossary/first-time-right. Accessed 4 July 2024