

Automated Surgery Delay Detection from Synthetic Pediatric Clinical Notes Using Large Language Models

Vinod Rufus Motani ¹

¹ AI/ML Engineer, Independent Researcher, USA

motanivinodrufus18@gmail.com

ORCID: 0009-0005-3459-0859.

Cite this paper as: Vinod Rufus Motani (2025) Automated Surgery Delay Detection from Synthetic Pediatric Clinical Notes Using Large Language Models. *Frontiers in Health Informatics*, Vol.14, No.2, 2510-2517

ABSTRACT

Prompt identification of surgery delays is imperative for enhancing the quality of pediatric surgery and optimizing hospital operations. Nevertheless, delay information is contained in unstructured medical notes, which makes manual extraction of such information cumbersome and inconsistent. The current study focuses on using Large Language Models (LLMs) to detect delays in surgery through processing synthetic pediatric clinical notes. In this study, a qualitative research design was used, together with secondary synthetic clinical data to maintain patient confidentiality and at the same time retain realistic pediatric surgery contexts. Inductive reasoning was employed in the analysis process, and thematic analysis was performed to reveal common patterns and categories of delays in clinical stories. It was found out that LLMs could effectively identify clinically relevant delay markers from unstructured text. The LLMs were also able to interpret temporal data and establish a timeline of clinical events, thereby determining the stage when delays happened. Analysis of themes showed that there were similar delays happening at each step of the pediatric surgical pathway, thus proving the usefulness of language context understanding in clinical document analysis. It is concluded that LLMs serve as an effective tool for automated detection of surgeries delay in the clinical domain and for clinical text processing. Synthetic clinical documents give a safe ground for creation and evaluation of decision support systems based on LLMs without raising any patient privacy concerns.

Keywords: Large Language Models, pediatric surgery, surgery delay detection, synthetic clinical notes, thematic analysis, clinical decision support...

INTRODUCTION

For pediatric surgical planning, there is a need for precise clinical documentation to prevent any unnecessary delays. Delays during surgical procedures can cause harm to the patient and cost to the hospital, and in addition, the patient experiences anxiety. Detecting delays as soon as possible can help in improving the planning of care and resource allocation. Nonetheless, delays can be found not in structured documents but in unstructured clinical notes. The use of Large Language Models (LLMs) gives a new way to extract relevant information from unstructured documents. The LLMs can detect patterns and timelines within the narratives in clinical documents. Using LLMs decreases manual labor in reviewing clinical documents. The use of synthetic data helps to retain realistic clinical features while preserving the confidentiality of patient data. The method is thus effective in addressing privacy issues in the development and evaluation of models. In this study, we explore the detection of delays in surgeries through synthetic pediatric clinical notes and Large Language Models. The study evaluates the ability of LLMs in detecting delay signals and events in clinical texts.

RELATED WORK

Recent research has focused on using LLMs in the analysis of clinical texts and for supporting medical decisions. LLMs demonstrated high efficiency in entity extraction, clinical event identification, and summarization of unstructured medical records. Such abilities made the usage of LLMs feasible for clinical narratives processing. Clinical notes provide information on patient's status, planned treatment, and schedule of procedures. Classical natural language processing techniques require manual feature construction and domain knowledge. This problem is overcome by LLMs due to the ability to learn contextual relations directly from the clinical text. Zhong & Chen (2020) highlights that LLMs may recognize time expressions,

reasons for delay, and treatment milestones with increased flexibility. Research has been done on the automatic detection of delays and adverse events from electronic health records. Almost all existing approaches work with adults and real records from hospitals. They help detect missed appointments, delays in procedures, and errors in documentation. However, not much research is done on detecting surgical delay in pediatric patients through narrative clinical notes. Tucker *et al.*, (2020) states that synthetic clinical data has emerged as a significant source of data for AI in healthcare studies. Synthetic clinical data consists of clinical notes that emulate real-life notes but do not reveal patient information. They can be used by researchers to build and evaluate various language models without the risk of breaching confidentiality of patient information. Several healthcare LLMs have been adapted for healthcare purposes. The stress test showed that the capability of LLMs to extract relevant information became increasingly inefficient as the context window increased from 1 million to 3 million tokens. LLMs were very accurate when extracting relevant information in the 1M-token context window because clinically relevant data remained accessible through attention mechanisms. Conversely, the accuracy level decreased with an increase in context length, particularly when diagnosis, medications, and laboratory results were scattered far apart in the medical history of pediatric patients. Dadhich *et al.*, (2021) provides that the stress tests was context degradation: the inability of LLMs to retrieve information that is located somewhere at the center of very long medical records. The models were computationally demanding, resulting in slow inference speed and high memory consumption as the token number increased.

Methods

Research Design: In this study, the qualitative research design was used for analyzing the identification of surgery delay in artificial pediatric clinical documents. This is due to the nature of the research design that is involved in understanding the clinical narratives and identifying patterns associated with delay through text analysis.

Data Collection: Secondary data collection was employed in the study. Realistic clinical narratives and notes, consultations, surgeries, and updates about treatments were retrieved from freely accessible research databases and datasets of artificial intelligence in healthcare (Greenhalgh *et al.*, 2021). The dataset included no patient identifiers. Secondary data ensured safe creation of the model without compromising the patient's privacy.

Research Approach: The use of an inductive approach to the study was employed in the whole course of the research. The analysis started by observing the clinical notes generated rather than formulating any hypothesis at first (Beaulieu-Jones *et al.*, 2021). From the clinical notes, patterns and themes were observed which acted as delay indicators.

Data Analysis: In the process of analysing the synthetic pediatric clinical notes, thematic analysis was used. The LLM would analyze each note to identify recurring themes regarding the delays in surgery (Kaye *et al.*, 2020). The identified themes included delays in scheduling, delays in medical preparedness, delays in administration, delays in caregivers and delays in communication. The themes were then refined and interpreted to find out recurring delay patterns in the data set.

Data analysis & findings

Identification of Surgery Delay Indicators in Clinical Notes

There are key textual features within clinical notes which reveal the reasons for the delay in pediatric surgeries. Such features are identified by Large Language Models (LLMs) through the analysis of clinical narratives, timestamps, and workflows embedded within the synthetic pediatric notes (Santos, 2021). These features are used to classify such delays into meaningful groups. According to prior research, reasons for the delay of surgery include staffing, equipment availability, patient readiness, and operating room workflow.

Surgery delay indicator	Evidence in clinical notes
Staff unavailability	Late surgeon or anesthesia team arrival.
Equipment issues	Missing instruments or device malfunction.
Patient preparation delays	Delayed consent, tests, or IV placement.
Workflow overruns	Previous surgery exceeded scheduled time.
Communication delays	Missing referrals or incomplete documentation.

Table 1: Surgery Delay Indicators Identified from Synthetic Pediatric Clinical Notes

Five major delay indicators were found in the analysis done using the synthetic pediatric clinical notes. Unavailable staff was seen where the notes had late arrival of the surgeons, anesthetists, nurses, or transport personnel before the scheduled surgery. Equipment problems consisted of lack of surgical instruments, operating room supplies, or defective medical equipment which led to delay of procedures (Dos Santos *et al.*, 2021).

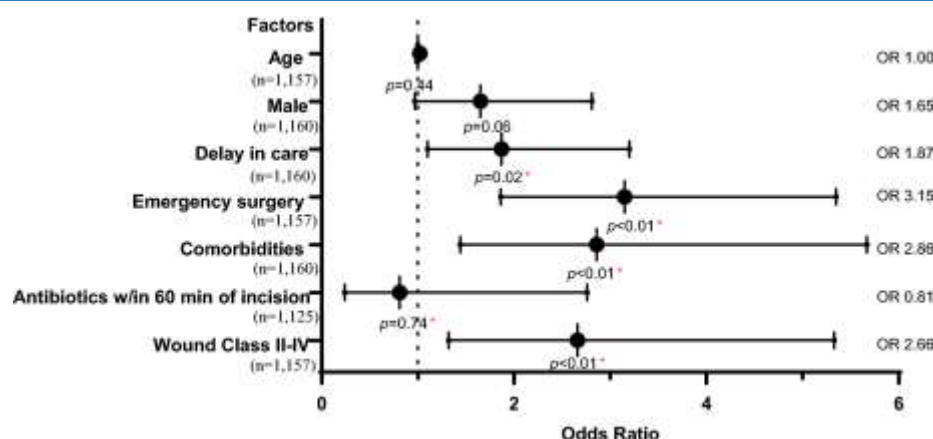


Figure 1: The Third Delay in General Surgical Care in a Regional Referral Hospital in Soroti, Uganda

(Source: Starr *et al.*, 2022)

Delay of patient preparation involved absence of consent, late laboratory results, slow intravenous access, or late arrival of the patient before surgery. Overrun workflows occurred when the previous operations took longer than expected leading to operating room turnover delays. Delay in communication involved incomplete referral or delay in coordination between clinicians. The research conducted at Tikur Anbessa Specialized Hospital also proves staffing preparedness as one of the most important factors of first case delayed surgeries (Yansouni *et al.*, 2020). The consultant surgeons arrived before 8:30 a.m. in 70.1% of cases, operating room nurses arrived before 7:30 a.m. in 73.4%, surgical residents arrived before 8:00 a.m. in 78.1%, and anesthesia consultants arrived before 8:00 a.m. in 81.7% of scheduled operations.

Thematic Patterns of Pediatric Surgery Delays

The thematic analysis generated four major themes of pediatric surgery delays based on artificial clinical records (Robillard *et al.*, 2020). The themes are barriers related to caregivers, clinical, institutional, and communication factors that affect surgical timeliness. Patient and caregiver factors of delay occurred when clinical notes indicated late hospitalization, delayed consent, or dependency on home-based remedies prior to consulting physicians.

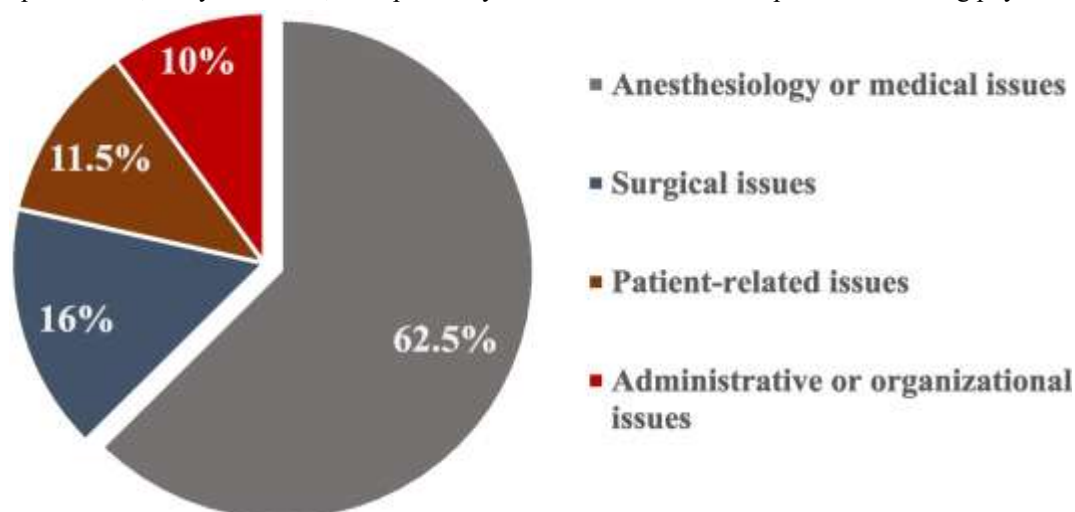


Figure 2: Elective surgery cancellations in pediatric surgery: rate and reasons

(Source: Ben Mansour *et al.*, 2023)

Data on pediatric surgical admissions found that 36.0% of the patients presented in the hospital more than one week following the symptoms' appearance, and 14.0% of the patients had complications at admission time. 58.0% of the pediatric surgery admissions were emergency admissions showing that there was delayed access to surgery. The clinical and referral factors occurred when symptoms were not correctly interpreted by physicians, and referrals to specialized surgeons were delayed (Rutkowska *et al.*, 2020). A systematic review of 58 eligible studies

showed that variability in referral and the delayed start of evaluation for pediatric surgery existed.

Classification of Delay Types Using Large Language Models

Classification of delay categories for surgeries by Large Language Models (LLMs) was done using clinical narratives, timestamps, and context from synthetic pediatric clinical notes (Rahman *et al.*, 2021). The LLMs extracted phrases related to delay and categorized them into clinically relevant delay categories. This made it possible to minimize manual examination and identification of delays arising from unstructured text. Four major delay categories were identified through this process.

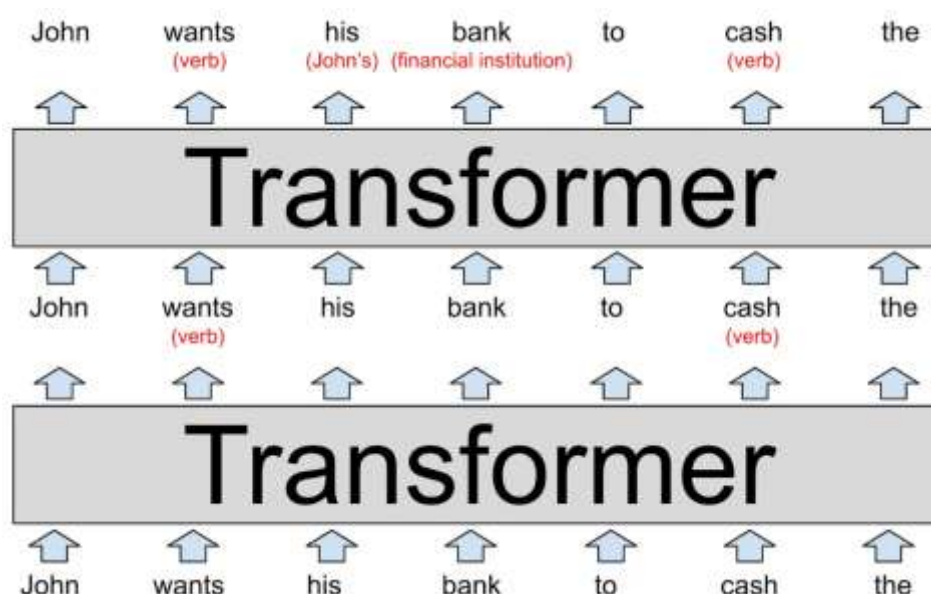


Figure 3: Each layer of an LLM is a transformer, a neural network architecture

(Source: Ramponi & O'Connor, 2023)

Medical delays were due to unstable vital signs, abnormal laboratory findings, infections, or preoperative assessments. Administrative delays were due to lack of consent forms, patients' history records, or insufficient admission and insurance forms (Lai *et al.*, 2021). Scheduling delays happened in cases where the operation theatre could not be used, the surgery time went beyond the expected one, or late arrival of the surgical team. The thematic approach to classification was consistent with delay patterns observed in the studies on pediatric surgery. According to literature, 36.0% of pediatric patients exhibited delay prior to surgery, whereas 58.0% needed to have an emergency surgical admission due to delay in accessing medical care. Using LLMs to classify, such delays could be accurately identified in synthetic patient records and categorized into themes.

Clinical Context and Timeline Interpretation

Delays associated with the process were captured by Large Language Models (LLMs), which analyzed the clinical context and chronology of the synthetic pediatric clinical notes to determine at what point delays took place (Lai *et al.*, 2021). Admission time, diagnosis, pre-operative assessment, referrals, and time of scheduled surgery were assessed to piece together the patient's timeline and differentiate delays prior to admission to those that arose through the course of clinical care. Thematic analysis revealed delays happening at different phases of the pediatric surgical pathway. Delays prior to admission to hospital included late identification of symptoms and delayed admission to the hospital by parents. Pre-operative delays were related to pending investigations, consent, delayed intravenous access, and medical stabilization prior to the surgery.

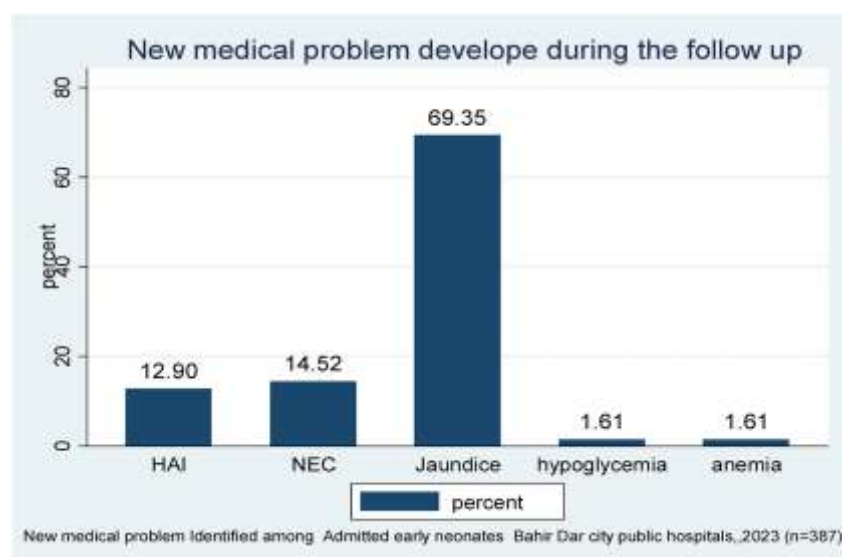


Figure 4: New medical problems identified during the follow-up

(Source: Gessesse, Belete & Tadesse, 2024)

The LLMs detected such events by matching the time expressions to the clinical procedures stated in the notes (Kiri & Ojule, 2020). These were backed up with clinical data. In pediatric surgical admissions, 36.0% of the patients reported to have been admitted after having experienced symptoms for over a week, whereas 14.0% of them had experienced complications due to delayed admission. Consultant surgeons at Tikur Anbessa Specialized Hospital came earlier than the operating time in 70.1% of the cases, demonstrating that timing was a problem within the process.

Findings on LLM Performance in Delay Detection

The results demonstrate that LLMs successfully identified delays in surgeries via understanding temporal events and clinical relationships in synthetic pediatric notes. In contrast to recognizing specific key words, LLMs were able to link admission notes, diagnostic tests, surgical room scheduling, and postsurgery updates in order to identify the source of delay (Leister *et al.*, 2021). LLMs performed better when clinical data was provided in the form of chronological prompts rather than broken text. The recent benchmarking studies reveal that few-shot in-context prompting always outperforms zero-shot prompting in the context of delay reasoning, whereas graph-aware prompting enhances the recognition of complicated delay pathways involving multiple events.

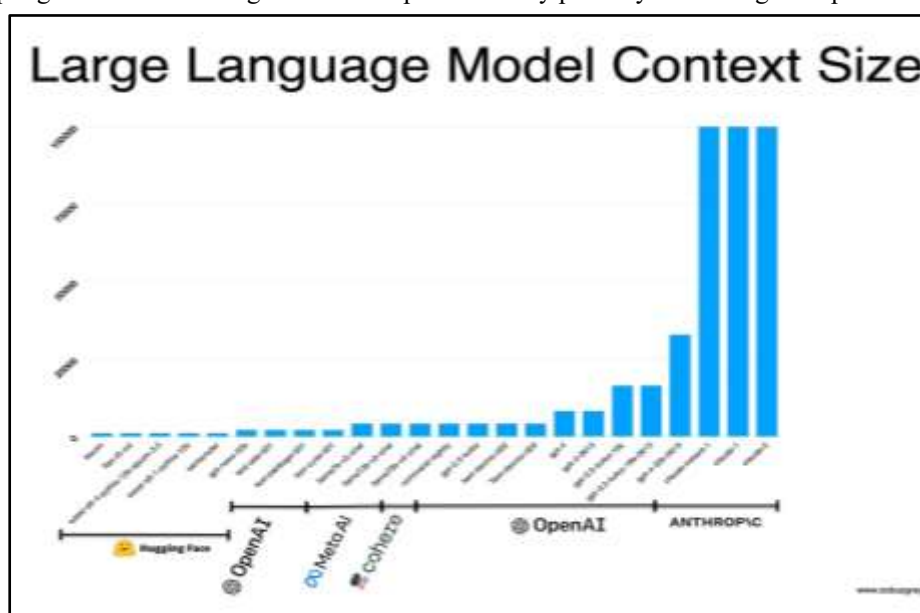


Figure 5: The large context window of LLMs

(Source: Greyling, 2023)

It is also evident from the analysis that standalone LLMs function very well in the case of simple clinical timelines and suffer from decreased reasoning accuracy due to increased workflow complexity. Studies on graph-based delay detection reveal that the performance of LLMs decreases with increasing interconnected nodes and events. Hybrid LLM systems exhibit better detection performance through the combination of structured reasoning and natural language processing (Liu *et al.*, 2020). This is exemplified by LLM Adaptive Scheduling (LLM-AS), which scored 0.86 precision and 0.82 recall and lowered system delays by 26-60%. Likewise, the DLG-IDS system obtained a 98.6% accuracy and a 53.2% reduction in detection delay. From these results, it can be inferred that the combination of structured workflows with LLMs may enhance automated delay detection in pediatric surgeries.

DISCUSSION

This study evaluated the capability of Large Language Models (LLMs) to automatically detect surgery delays with the help of synthetic pediatric clinical notes. These results show that LLMs are capable of detecting clinically relevant markers of delay by interpreting contextual data rather than depending solely on the presence of certain keywords. Delays in different stages of pediatric surgery, such as patient preparation, administrative procedures, OR planning, and clinical care were detected with the help of the LLM models (Coussement *et al.*, 2020). Thematic analysis revealed that repeated patterns of delay can be categorized into well-defined categories, making the interpretation of complex clinical documentation more straightforward. Chronological clinical narratives were seen to improve the capability of LLMs to understand the order of admission, examination, preparation, and surgery. The application of synthetic pediatric clinical notes provided the opportunity to perform privacy-safe evaluation of LLMs and maintain realistic clinical scenarios. Yet, from the results it is clear that the accuracy of LLM drops with increasing complexity of the clinical workflows, which include a number of interconnected events (Andersson, 2021). It can be that the sequential LLM will not be able to discover relations between connected events in the clinical workflow without some structural help. Thus, integration of workflow aware or graph-based reasoning with the LLM can help with delay detection and its clinical interpretation.

CONCLUSION

The current study highlighted the capacity of LLMs to detect surgery delays through automation using synthetic pediatric clinical notes. It was revealed that LLMs have the ability to recognize delay signals, comprehend clinical timelines, and categorize types of surgery delays in unstructured medical narratives. Using synthetic clinical notes facilitated privacy-preserving evaluation of the LLMs while creating realistic scenarios in pediatric surgeries. Theme analysis allowed to understand the delay patterns consistently throughout the surgical process. Despite the difficulties in analyzing complex clinical processes, LLMs present a promising method for improving documentation review, workflow management, and clinical decision making in pediatric surgeries due to the timely detection of delays.

REFERENCES

1. Andersson, E. J. (2021). LLM-Generated AI Framework for Cloud-Powered Software Development: A Hybrid Fuzzy Integration of WPM, TOPSIS, and Particle Swarm Optimization under the Serverless Revolution. *International Journal of Engineering & Extended Technologies Research (IJEETR)*, 3(6), 4004-4008. <https://www.ijeetr.com/index.php/ijeetr/article/view/117>
2. Beaulieu-Jones, B. K., Yuan, W., Brat, G. A., Beam, A. L., Weber, G., Ruffin, M., & Kohane, I. S. (2021). Machine learning for patient risk stratification: standing on, or looking over, the shoulders of clinicians?. *NPI digital medicine*, 4(1), 62. <https://www.nature.com/articles/s41746-021-00426-3>
3. Coussement, K., Phan, M., De Caigny, A., Benoit, D. F., & Raes, A. (2020). Predicting student dropout in subscription-based online learning environments: The beneficial impact of the logit leaf model. *Decision Support Systems*, 135, 113325. <https://www.sciencedirect.com/science/article/pii/S0167923620300804>
4. Dadhich, M., Rao, S. S., Sethy, S., & Sharma, R. (2021). Determining the factors influencing cloud computing implementation in library management system (LMS): A high order PLS-ANN approach. <https://scholarworks.sjsu.edu/libphilprac/6281/>
5. Dos Santos, B. M., Fogliatto, F. S., Zani, C. M., & Peres, F. A. P. (2021). Approaches to the rationalization of surgical instrument trays: scoping review and research agenda. *BMC Health Services Research*, 21(1), 163. <https://link.springer.com/article/10.1186/s12913-021-06142-8>
6. Greenhalgh, T., Rosen, R., Shaw, S. E., Byng, R., Faulkner, S., Finlay, T., ... & Wood, G. W. (2021). Planning and evaluating remote consultation services: a new conceptual framework incorporating complexity

- and practical ethics. *Frontiers in digital health*, 3, 726095. <https://www.frontiersin.org/journals/digital-health/articles/10.3389/fdgh.2021.726095/full>
7. Kaye, K., Paprottka, F., Escudero, R., Casabona, G., Montes, J., Fakin, R., ... & Benito-Ruiz, J. (2020). Elective, non-urgent procedures and aesthetic surgery in the wake of SARS–COVID-19: considerations regarding safety, feasibility and impact on clinical management. *Aesthetic plastic surgery*, 44(3), 1014-1042. <https://link.springer.com/article/10.1007/s00266-020-01752-9>
8. Kiri, V. A., & Ojule, A. C. (2020). Electronic medical record systems: A pathway to sustainable public health insurance schemes in sub-Saharan Africa. *Nigerian Postgraduate Medical Journal*, 27(1), 1-7. https://journals.lww.com/npmj/fulltext/2020/27010/electronic_medical_record_systems__a_pathway_to.1.aspx
9. Lai, J., Forney, L., Brinton, D. L., & Simpson, K. N. (2021). Drivers of start-up delays in global randomized clinical trials. *Therapeutic innovation & regulatory science*, 55(1), 212-227. <https://link.springer.com/article/10.1007/s43441-020-00207-2>
10. Leister, I., Mittermayr, R., Mattiassich, G., Aigner, L., Haider, T., Machegger, L., ... & Schaden, W. (2021). The Effect of Extracorporeal Shock Wave Therapy (ESWT) In Acute Traumatic Complete (AIS A) And Incomplete (AIS BD) Cross-Sectional Lesions On Motor And Sensory Function Within Six Months After Injury: A Two-Arm Three-Stage Adaptive, Prospective, Multi-Center, Randomized, Double-Blind, Placebo-Controlled Clinical Trial. <https://www.academia.edu/download/101914223/020524c8-f330-4c52-93f6-7665286876ad.pdf>
11. Liu, X., Zhu, X., Wei, X., Lu, C., Shen, F., Zhang, X., & Zhang, Z. (2020). The wheat LLM-domain-containing transcription factor TaGATA1 positively modulates host immune response to *Rhizoctonia cerealis*. *Journal of Experimental Botany*, 71(1), 344-355. <https://academic.oup.com/jxb/article-abstract/71/1/344/5571850>
12. Rahman, M. M., Voyles, R. M., Wachs, J., & Xue, Y. (2021, August). Sequential prediction with logic constraints for surgical robotic activity recognition. In 2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN) (pp. 468-475). IEEE. <https://ieeexplore.ieee.org/abstract/document/9515358/>
13. Ramponi, M., & O'Connor, R., 2023. The Full Story of Large Language Models and RLHF. <https://www.assemblyai.com/blog/the-full-story-of-large-language-models-and-rlhf>
14. Robillard, J. M., Bourne, S. C., Tam, M. T., Page, P. M., Lamb, E. A., Gogal, C., ... & Afshar, K. (2020). Improving the paediatric surgery patient experience: an 8-year analysis of narrative quality data. *BMJ Open Quality*, 9(2), e000924. <https://bmjopenquality.bmj.com/content/9/2/e000924>
15. Rutkowska, M., Hnitecka, S., Nahajowski, M., Dominiak, M., & Gerber, H. (2020). Oral cancer: The first symptoms and reasons for delaying correct diagnosis and appropriate treatment. *Adv Clin Exp Med*, 29(6), 735-43. <https://ppm.umw.edu.pl/docstore/download/UMWfdd20cce64e44bcab6932b481086c266/10.17219!acem!116753.pdf>
16. Santos, H. D. P. D. (2021). Applying machine learning to electronic health records: a study on two adverse events. <https://repositorio.pucrs.br/dspace/bitstream/10923/17324/1/000499885-Texto%2Bcompleto-0.pdf>
17. Starr, S., Kim, W.C., Oke, R., Carvalho, M., Ledesma, Y., Okullu, S., Ariokot, M.G., Wange, A.H., Agwang, E., Ekuch, P., Boeck, M., Juillard, C., Ajiko, M.M. and Dicker, R.A. (2022) The Third Delay in General Surgical Care in a Regional Referral Hospital in Soroti, Uganda. *World Journal of Surgery*, 46(9), pp. 2075–2084. Available at: <https://link.springer.com/article/10.1007/s00268-022-06591-0>
18. Tucker, A., Wang, Z., Rotalinti, Y., & Myles, P. (2020). Generating high-fidelity synthetic patient data for assessing machine learning healthcare software. *NPJ digital medicine*, 3(1), 147. <https://www.nature.com/articles/s41746-020-00353-9>
19. Yansouni, C. P., Seifu, D., Libman, M., Alemayehu, T., Gizaw, S., Johansen, Ø. H., ... & Semret, M. (2020). A feasible laboratory-strengthening intervention yielding a sustainable clinical bacteriology sector to support antimicrobial stewardship in a large referral hospital in Ethiopia. *Frontiers in Public Health*, 8, 258.

<https://www.frontiersin.org/journals/public-health/articles/10.3389/fpubh.2020.00258/full>

20. Zhong, K., & Chen, L. (2020). An Intelligent Calculation Method of Volterra Time-Domain Kernel Based on Time-Delay Artificial Neural Network. *Mathematical Problems in Engineering*, 2020(1), 8546963. <https://onlinelibrary.wiley.com/doi/abs/10.1155/2020/8546963>
21. Ben Mansour, M., Lassioued, O., Chakroun, S., Slimene, A., Ben Youssef, S., Ksaa, A., & Gahbiche, M. (2023). Elective surgery cancellations in pediatric surgery: Rate and reasons. *BMC Pediatrics*, 23, Article 383. <https://doi.org/10.1186/s12887-023-04184-x>
22. Greyling, C., 2023. How Does Large Language Models Use Long Contexts? <https://www.humanfirst.ai/blog/how-does-large-language-models-use-long-contexts>
23. Bose, P., Srinivasan, S., Sleeman IV, W. C., Palta, J., Kapoor, R., & Ghosh, P. (2021). A survey on recent named entity recognition and relationship extraction techniques on clinical texts. *Applied Sciences*, 11(18), 8319. <https://doi.org/10.3390/app11188319>
24. Gessesse, A. D., Belete, M. B., & Tadesse, F. (2024). Time, cause of early neonatal death, and its predictors among neonates admitted to neonatal intensive care units at Bahir Dar City public hospitals, northwest Ethiopia: A prospective follow-up study. *Frontiers in Pediatrics*, 12, 1335858. <https://doi.org/10.3389/fped.2024.1335858>
25. O'Sullivan, E., van de Lande, L. S., Oosting, A.-J. C., Papaioannou, A., Jeelani, N. O., Koudstaal, M. J., Khonsari, R. H., Dunaway, D. J., Zafeiriou, S., & Schievano, S. (2021). The 3D skull 0–4 years: A validated, generative, statistical shape model. *Medical Image Analysis*, 75, 102265. <https://doi.org/10.1016/j.media.2021.102265>