Intelligent Human Systems Integration (IHSI 2025), Vol. 160, 2025, 576–584
https://doi.org/10.54941/ahfe1005863

Optimizing Healthcare Efficiency With Local Large Language Models

Ivan Lorencin, Nikola Tanković, and Darko Etinger

Juraj Dobrila University of Pula, Faculty of Informatics, Croatia

ABSTRACT

Large Language Models (LLMs) are increasingly recognized for their potential to alleviate administrative burdens in healthcare, enabling medical professionals to focus more on patient care rather than time-consuming administrative tasks. This paper explores how local LLMs can support healthcare settings by automating and streamlining routine administrative duties, improving workflow efficiency, and ultimately enhancing patient care. One of the key applications of LLMs is in the management of medical documentation. Healthcare professionals often spend significant time on tasks such as transcribing patient notes, updating medical records, and completing forms. By using LLMs, these processes can be automated or simplified. The models can transcribe and structure patient interactions in real-time, generate diagnostic summaries, and update electronic health records (EHRs) based on structured inputs, reducing the time healthcare providers spend on paperwork. This not only saves valuable time but also minimizes the risk of errors associated with manual data entry. Another area where LLMs can be beneficial is in appointment scheduling and patient communication. LLMs can be integrated into practice management systems to manage appointments, send reminders, and handle patient inquiries. By processing natural language requests, these models can schedule or reschedule appointments, direct patients to the appropriate specialists, and provide answers to frequently asked questions. This reduces the administrative workload for healthcare staff, allowing them to focus on more critical tasks and improving overall clinic efficiency. In addition, LLMs can assist with billing and insurance processing. By automatically extracting relevant information from patient records and claims, LLMs can generate billing codes, verify insurance coverage, and ensure that all necessary documentation is submitted. This reduces the administrative burden on healthcare providers and billing departments, streamlining the reimbursement process and minimizing errors in insurance claims. Local LLMs also aid in regulatory compliance by automatically ensuring that healthcare institutions adhere to relevant legal requirements, such as patient consent and privacy regulations. By continuously monitoring the creation and modification of medical records, these models can flag potential issues related to compliance and generate alerts, ensuring that healthcare providers remain in line with regulations such as GDPR. This proactive approach to compliance reduces the risk of legal liabilities and minimizes the time spent on manual checks. In conclusion, by automating tasks such as documentation, scheduling, billing, and regulatory compliance, local LLMs can improve workflow efficiency, reduce human error, and enhance overall productivity in healthcare settings. As the technology evolves, local LLMs have the potential to significantly transform healthcare operations, allowing medical professionals to focus on what matters most: their patients.

Keywords: Data privacy, Healthcare system, Llama 3.2, Local LLMS, NotebookLM

© 2025. Published by AHFE Open Access. All rights reserved.

INTRODUCTION

Healthcare systems worldwide are under increasing pressure to deliver high-quality care while managing growing administrative workloads. This challenge is particularly pronounced in systems like Croatia's, where systemic inefficiencies and resource allocation issues hamper performance. In 2018, Croatia's overall healthcare efficiency was just 57%, ranking it among the least efficient in the European Union. Despite achieving cost-efficiency, systemic inefficiencies—such as the low transformation of inputs into outcomes—persist. Compounding this is an aging population, limited investment in preventive care, and high rates of preventable diseases such as ischemic heart disease and lung cancer. These factors necessitate innovative solutions to optimize healthcare operations without requiring significant additional resources.

The integration of artificial intelligence, particularly large language models (LLMs), offers a transformative opportunity to address these challenges. LLMs enable automation in key areas such as medical documentation, appointment scheduling, and billing, significantly reducing administrative burdens. These systems are particularly relevant in the Croatian context, where digital health initiatives like "Portal Zdravlja" have already demonstrated the potential of technology to enhance efficiency and accessibility. Digital solutions can play a vital role in reallocating resources toward preventive care and improving overall health outcomes.

A critical aspect of deploying LLMs in healthcare is ensuring data privacy and compliance with stringent regulations such as GDPR and HIPAA. We hypothesize that local deployment of LLMs offers superior data privacy protection compared to cloud-based systems. This is especially critical in healthcare, where sensitive patient information must remain secure and accessible only to authorized personnel. By processing data locally, institutions mitigate risks associated with external data transmission and enhance patient trust through robust privacy safeguards. In addition to addressing administrative inefficiencies, LLMs align well with ongoing digital transformation efforts in Croatia. For example, initiatives like the European Health Data Space (EHDS) aim to standardize and enhance health data sharing across the EU, presenting opportunities for countries like Croatia to improve resource utilization and coordination of care. LLMs, deployed locally or in hybrid models, can complement such frameworks, enabling efficient data processing while maintaining compliance with privacy standards.

This paper explores the role of LLMs in optimizing healthcare operations, with a particular focus on local deployments tailored to the needs of community health centers and systems like Croatia's. By leveraging these technologies, healthcare providers can achieve greater efficiency, enhance patient care, and meet the growing demands of modern healthcare delivery while safeguarding sensitive patient data.

EFFICIENCY OF THE CROATIAN HEALTHCARE SYSTEM

The Croatian healthcare system faces significant challenges in resource allocation and utilization. In 2018, overall efficiency was measured at 57%, placing Croatia among the least efficient EU countries. While cost-efficiency achieved a perfect score of 100%, systemic efficiency—reflecting the conversion of intermediate inputs into health outcomes—was only 48%, as presented in Figure 1 (Lacko et al., 2022). As illustrated in the chart, Croatia lags behind neighboring countries in the region in healthcare efficiency. Despite efforts to optimize resources, significant systemic inefficiencies remain, highlighting the need for further reforms and investments. This suggests that Croatia could enhance healthcare outcomes by addressing internal inefficiencies without increasing resource input (Buljan & Šimović, 2022).

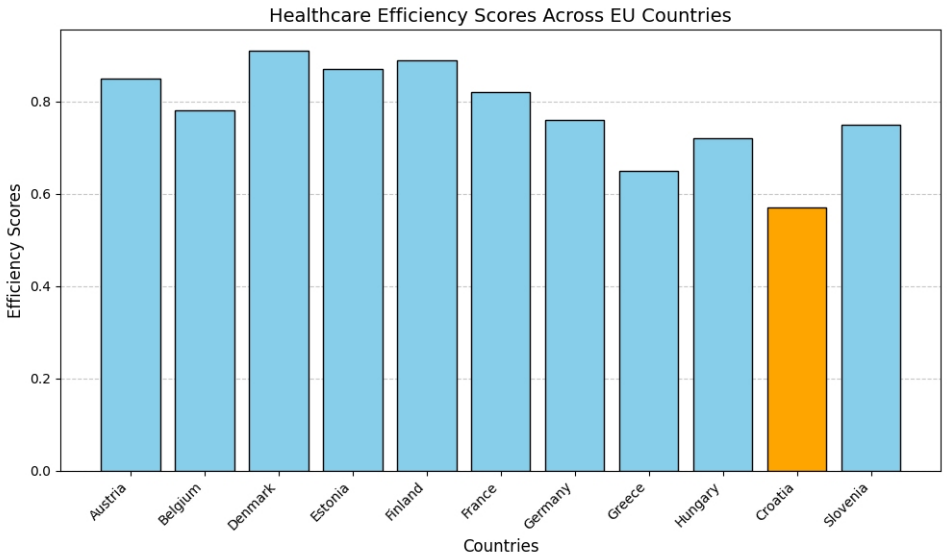

Figure 1: Healthcare efficiency scores across EU countries.

Health indicators further highlight the system's underperformance. In 2020, life expectancy in Croatia was 77.8 years, nearly three years below the EU average. High mortality rates from preventable causes, such as ischemic heart disease and lung cancer, underscore the need for systemic reforms. Notably, only 3% of total healthcare expenditure is allocated to preventive care, slightly above the EU average of 2.9%. Healthcare spending in Croatia accounts for 6.9% of GDP, below the EU average of 8.3%. Per capita expenditure is approximately €1,361, significantly less than the EU average of €2,572. Despite substantial public funding—82.8% from mandatory insurance and budget transfers—resource allocation remains inefficient, with disproportionate spending on pharmaceuticals and minimal investment in preventive care. Digital transformation presents a significant opportunity to address these challenges. The COVID-19 pandemic accelerated the adoption

of telemedicine and other digital health innovations, underscoring the potential of digital solutions to improve healthcare accessibility and efficiency (Funda et al., 2022).

However, the pandemic also exposed existing inequalities in healthcare access, emphasizing the need for digital health solutions to be tailored to diverse patient needs and capabilities. Technology should enhance, not replace, human interaction in healthcare, ensuring quality care for all patients regardless of socioeconomic status or geographic location.

One of the greatest challenges in achieving this goal is the shortage of healthcare personnel (Mustajbegović, 2023), which is particularly pronounced in rural areas of Croatia, as well as in coastal regions due to the high cost of living, making it even more difficult to attract and retain skilled professionals. Croatia's healthcare system has substantial potential for efficiency improvements through digital transformation. By addressing systemic inefficiencies, reallocating resources towards preventive care, and embracing digital health solutions, Croatia can achieve better health outcomes and improve its standing relative to other EU countries. These reforms are essential for ensuring the sustainability and effectiveness of the healthcare system amid demographic changes and increasing healthcare demands. Healthcare professionals dedicate substantial time to tasks such as transcribing patient notes, updating medical records, and completing forms. While crucial, these tasks detract from time available for direct patient care. Integrating large language models (LLMs) offers transformative potential by automating transcription and structuring patient interactions. LLMs enable instant updates to electronic health records (EHRs), generating accurate summaries from structured inputs, which minimizes manual effort and errors. This automation not only enhances accuracy by reducing human input errors but also saves considerable time, allowing healthcare providers to prioritize patient interactions.

The implementation of automation is further empowered by advancements in digital health applications. Systems such as the Croatian "Portal Zdravlja" demonstrate the impact of digitized patient data management, where automation streamlines access to laboratory results, medical history, and prescriptions (Gvozdanović, 2024). This integration complements LLM-driven transcription by providing a unified platform for all healthcare data, ensuring consistency and security. Furthermore, as seen in Croatia's e-health initiatives, robust data protection protocols bolster confidence in digital systems, ensuring patient privacy and compliance with regulatory standards.

Efficient scheduling and communication remain pivotal to healthcare operations. LLMs seamlessly integrate with practice management systems, automating scheduling, rescheduling, and confirmations through natural language queries. Chatbots powered by LLMs efficiently address frequently asked questions, guide patients to appropriate services, and send reminders, thereby reducing administrative workload and improving patient satisfaction through prompt and accurate responses. These benefits align with digital health innovations like telemedicine platforms, which have gained prominence during the COVID-19 pandemic. For instance, Croatia's rapid adoption of mobile applications such as "Andrija" underscores

the synergy between AI and telecommunication technologies in patient engagement (Petricevic & Mustic, 2023). These platforms facilitate on-demand consultations, reducing the burden on administrative staff while enhancing patient accessibility. Healthcare billing and insurance processing are often complex, leading to delays and errors that impact both providers and patients. LLMs address these challenges by automating the extraction of relevant billing information, verifying patient eligibility, and ensuring compliance with insurance coverage requirements. By streamlining claim submissions and reducing manual intervention, these technologies minimize administrative overhead and accelerate reimbursement processes (Kang et al., 2024). The role of digitalization extends beyond automation to the centralization of billing and insurance data. Croatian initiatives in electronic health records and integrated information systems have shown how central databases can facilitate smoother interactions between patients, insurers, and healthcare providers. By leveraging LLMs alongside established digital platforms, healthcare systems can enhance operational efficiency and reduce discrepancies in billing. To evaluate the potential of local LLM solutions in healthcare, this study focuses on comparing their performance across various critical tasks. Key areas of evaluation include the ability to retrieve and summarize patient-specific information, provide accurate and relevant responses, and ensure coverage of essential details. By analyzing these capabilities through structured test buckets and metrics such as F1 Score, BLEU Score, Recall@K, and Key Information Coverage (KIC), the study aims to highlight the strengths and areas for improvement of solutions like NotebookLM and Llama 3.2-powered AnythingLM.

COMPREHENSIVE PATIENT MANAGEMENT WITH DUAL-RECORD SYSTEMS

Jane Doe is a synthetic patient created for the purpose of testing and evaluating healthcare management systems. Her case is documented in two separate records: a Clinical Record and an Administrative Record, each serving distinct purposes in her care process. Detailed patient information is provided in Table 1.

Table 1. Details about synthetic patient Jane Doe.

Record Type	Details
Diagnosis	High-grade serous carcinoma of the left ovary, FIGO stage IIIC
Clinical Record - Symptoms	Persistent abdominal pain, bloating, and weight loss
Clinical Record - Diagnosis	Confirmed diagnosis through imaging and biopsy results
Clinical Record - Treatment Plans	Chemotherapy regimen, surgical planning, and supportive care
Clinical Record - Progress Notes	Regular follow-ups showing partial response to treatment
Administrative Record - Insurance	Insurance details for claims processing and coverage verification
Administrative Record - Logistical Aspects	Coordination of appointments, transportation assistance, and support services

The Clinical Record provides comprehensive medical details about Jane. Records illustrate the integration of clinical and administrative workflows essential for comprehensive patient management. This dual-record system

also provides a robust foundation for evaluating how LLMs can optimize healthcare operations by automating tasks and improving access to critical patient information. To evaluate the performance of NotebookLM and Local deployment, a diverse set of questions was designed, each targeting a specific category of information. These questions cover key aspects of a patient's medical case, ranging from insurance details to treatment plans. The test buckets, along with sample questions, are outlined below:

- What kind of insurance coverage does the patient have?
- What were the results of the biopsy?
- What did the pelvic ultrasound reveal?
- What is the patient's medical history?
- When is the surgical consultation scheduled?
- What symptoms did the patient report?
- What is the patient's family medical history?
- What did the CA-125 test indicate?
- What was the finding on the MRI?
- What is the planned treatment for the patient?

NotebookLM

NotebookLM, developed by Google, represents a transformative tool for healthcare professionals seeking to streamline data management and decision-making. By integrating Large Language Model (LLM) functionalities into a notebook-like interface, NotebookLM allows for seamless querying and organization of complex medical and administrative data. This capability is particularly beneficial for healthcare environments where efficient access to information is critical for both administrative and clinical workflows (Mehta et al., 2024). One of the primary advantages of NotebookLM is its ability to facilitate advanced information retrieval. Healthcare professionals can use NotebookLM to interact with structured and unstructured data, including patient records, medical research, and policy documents. NotebookLM also supports decision-making processes by summarizing key findings from research papers, clinical guidelines, or patient histories. This feature is particularly valuable in multidisciplinary teams, enabling collaboration and ensuring that all stakeholders have access to the most relevant information. When NotebookLM is tested using synthetic patient and test questions, the results presented in Table 2 are achieved.

Table 2. Metrics per questions using NotebookLM.

Question	F1 Score	BLEU Score	Recall@K	KIC
Q1	0.5614	0.2646	0.0909	0.7273
Q2	0.3243	0.1501	0.1579	0.6316
Q3	0.5667	0.1194	0.1364	0.7727
Q4	0.2667	0.0687	0.0	0.4706
Q5	0.2759	0.1029	0.0	0.8
Q6	0.5714	0.1448	0.1429	0.8571

(Continued)

Table 2. Continued

Question	F1 Score	BLEU Score	Recall@K	KIC
Q7	0.4333	0.1558	0.87	0.5652
Q8	0.2466	0.0308	0.0	0.6429
Q9	0.5909	0.2612	0.1875	0.8125
Q10	0.3913	0.1811	0.1875	0.5625

ETHICAL CONSIDERATIONS IN USING PATIENT DATA WITH LARGE LANGUAGE MODELS

The integration of Large Language Models (LLMs) in healthcare brings forward both opportunities and ethical challenges, particularly concerning the use of sensitive patient data (Mirzaei et al., 2024). The following paragraph outlines the key ethical considerations that must guide this integration: The use of patient data in LLMs must prioritize privacy and data security. A fundamental ethical requirement is obtaining informed consent from patients. This includes transparently informing patients about how their data will be used, whether processed by LLMs, and the safeguards in place. Explicit consent ensures that patients remain in control of their data and builds trust in the technology (Burkhardt et al., 2023). Ethical use of LLMs must align with established regulations and guidelines. This includes maintaining detailed audit logs, implementing strict access controls, and ensuring compliance with regional laws on data protection. These measures not only prevent legal repercussions but also uphold ethical standards in patient care.

LOCAL DEPLOYMENT OF LLMS

Deploying LLMs locally at community health centers offers a unique opportunity to enhance care while addressing privacy and infrastructure challenges. Local health centers often lack the IT resources of larger hospitals, making cloud-free or hybrid LLM solutions more appealing. Tools like LM Studio and AnythingLM allow for cost-effective deployments that can run on existing or affordable hardware, minimizing the need for extensive upgrades. Local data processing ensures that sensitive patient information is not transmitted to external servers, significantly reducing privacy risks. Furthermore, LLMs can be tailored to specific community health center workflows, such as managing high patient loads with minimal staff, thus enhancing operational efficiency. When RAG pipeline designed with Llama 3.2 using LM Studio and AnythingLM is tested using synthetic patient and test questions, the results presented in Table 3 are achieved. Despite advantages, locally deployed LLM shows lower performances, when compared with off-the-shelf solutions like NotebookLM.

Table 3. Metrics per questions using locally deployed LLama 3.2.

Question	F1 Score	BLEU Score	Recall@K	KIC
Q1	0.6222	0.2884	0.1364	0.6364
Q2	0.5714	0.3882	0.1579	0.5263
Q3	0.1765	0.0104	0.1364	0.1364
Q4	0.2128	0.0326	0.1176	0.2941
Q5	0.4706	0.0746	0.0	0.8
Q6	0.4828	0.1467	0.2143	0.5
Q7	0.5455	0.2585	0.1304	0.5217
Q8	0.2308	0.0896	0.2143	0.2143
Q9	0.0625	0.0145	0.0625	0.0625
Q10	0.2333	0.0254	0.0625	0.4375

PERFORMANCE COMPARISON

The final comparison chart showcases the average performance of NotebookLM and Llama 3.2 across key evaluation metrics: F1 Score, BLEU Score, Recall@K, and Key Information Coverage. NotebookLM consistently outperforms Llama 3.2 in terms of information coverage and recall, demonstrating its precision in capturing key elements of reference answers, as can be seen in Figure 2. Conversely, Llama 3.2 shows strength in BLEU Score, indicating a better fluency and linguistic similarity in generated responses. The results underscore the complementary capabilities of the two models, with NotebookLM excelling in accuracy and detail, while Llama 3.2 offers more natural language generation.

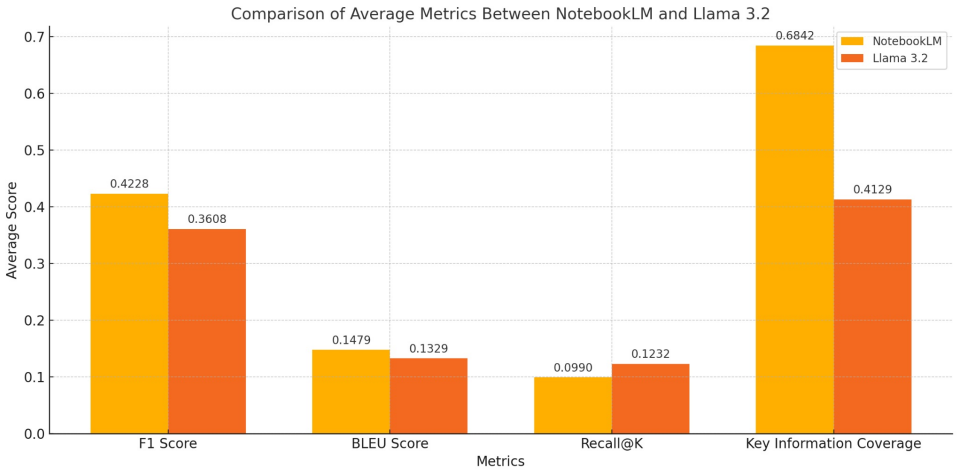

Figure 2: Comparison of achieved results.

CONCLUSION

Local LLMs represent a transformative opportunity for healthcare systems, automating routine tasks, enhancing efficiency, and maintaining high

standards of patient care and data privacy. Tools like NotebookLM and Llama 3.2 demonstrate the potential of local models, with NotebookLM showing consistent strength across multiple metrics. However, results also highlight the importance of considering a multi-model approach, where different LLMs are tailored to specific tasks, leveraging their unique strengths. Additionally, training new, domain-specific models could further optimize performance, addressing the nuanced demands of healthcare operations. By strategically integrating multiple local LLMs and developing bespoke models, healthcare providers can ensure a sustainable, efficient, and patient-centric future.

ACKNOWLEDGMENT

This research is (partly) supported by “European Digital Innovation Hub Adriatic Croatia (EDIH Adria) (project no. 101083838)” under the European Commission’s Digital Europe Programme, SPIN project "INFOBIP Konverzacijski Order Management (IP.1.1.03.0120)”, SPIN project “Projektiranje i razvoj nove generacije laboratorijskog informacijskog sustava (iLIS)” (IP.1.1.03.0158), and the FIPU project “Sustav za modeliranje i provedbu poslovnih procesa u heterogenom i decentraliziranom računalnom sustavu”.

REFERENCES

Buljan, A., & Šimović, H. (2022). Učinkovitost hrvatskog zdravstvenog sustava-usporedba sa zemljama Europske unije. Revija za socijalnu politiku, 29(3), 321–354.

Burkhardt, G. et al. (2023). Privacy behaviour: A model for online informed consent. Journal of business ethics, 186(1), 237–255.

Funda, D., et al. (2022). Digitalna transformacija zdravstvenog sustava. 23. medunarodni simpozij o kvaliteti kvaliteta-jucer, danas, sutra, 674, 601–610.

Gvozdanović, K. (2024). eKarton i Portal zdravlja u Republici Hrvatskoj. Bilten Hrvatskog društva za medicinsku informatiku, 30(1), 48–51.

Kang, I., et al. (2024). Using Large Language Models for Generating Smart Contracts for Health Insurance from Textual Policies. In AI for Health Equity and Fairness: Leveraging AI to Address Social Determinants of Health (pp. 129–146). Cham: Springer Nature Switzerland.

Lacko, R., et al. (2022). Efficiency and productivity differences in healthcare systems: The case of the European Union. International Journal of Environmental Research and Public Health, 20(1), 178.

Mehta, N., et al. (2024). Pedagogy and generative artificial intelligence: Applying the PICRAT model to Google NotebookLM. Medical Teacher, 1–3.

Mirzaei, T., et al. (2024). Clinician voices on ethics of LLM integration in healthcare: A thematic analysis of ethical concerns and implications. BMC Medical Informatics and Decision Making, 24(1), 250.

Mustajbegović, J. (2023). Očuvanje zdravlja i radne sposobnosti osoblja u djelatnosti zdravstva kao ključni čimbenik kvalitete zdravstvene zaštite. Liječnički vjesnik, 145(11–12), 393–397.

Petricevic, S., & Mustic, D. (2023). Research on Media Presentation and Public Reaction to the First Health Digital Assistant in Croatia. The Eurasia Proceedings of Science Technology Engineering and Mathematics, 23, 151–164.