IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

The Impacts of Artificial Intelligence on Reducing Surgical Errors in Heart Surgeries

Yonas Getenet Bekele

Menelik II Secondary School, King George VI St, Arat Kilo, Addis Ababa, Ethiopia; yonasgetenetbekele@gmail.com

ABSTRACT: Of all surgeries, heart surgeries are considered to pose a higher risk of surgical errors due to the intricacies of the heart's anatomical structures, surgical complexities, the critical nature of procedures, and stress on the surgical team. In this paper, considering the significant impact of artificial intelligence (AI) on enhancing patient outcomes in healthcare, we explore AI's potential in reducing errors in heart surgeries. To that end, we reviewed recent papers on the intersection of artificial intelligence and heart surgeries. We analyzed their efficiency in reducing surgical errors and their promise to mitigate the risks involved in cardiac surgeries. Large language-based surgical assistants, such as CARES Copilot 1.0 and ChatGPT, are cost-effective solutions for underprivileged areas with high demand for cardiac surgeries because they enhance surgical outcomes, shorten ICU and hospital stays, and thus give opportunities to other patients awaiting surgery. At the same time, despite reducing surgical errors and improving patient outcomes, AI has some limitations, including hallucinations, biases in training data, threats of job displacement, and high deployment costs, especially for robotic surgeries and fine-tuning large language models, all of which remain significant barriers to AI's adoption, requiring further research.

KEYWORDS: Artificial Intelligence (AI), surgical errors, robotic surgery, machine learning, cardiac surgery, CARES Copilot 1.0, ChatGPT, AI hallucinations.

Introduction

Heart surgery, also known as cardiac surgery, is a medical procedure that doctors use to treat diseases affecting the heart and its blood vessels. The surgeries include repairing defects in the heart, repairing damaged blood vessels, replacing heart valves, and removing blood clots from veins.[1] Access to heart surgeries is costly and not easily accessible globally. Currently, 6 billion people worldwide do not have access to cardiac surgical care.[2] This is even worse for middle- and low-income countries. In Ethiopia, there are only five cardiac surgeons serving 120 million people, and more than 15,000 patients are on the waitlist for surgery.[3] AI systems can play a potential role in improving the accessibility of cardiac surgical care through integration with the entire treatment process, allowing surgeons to perform intricate surgeries in less time with high precision and accuracy.

Medical errors can occur in several forms, including surgical, diagnosis, medication, equipment, patient identification, testing, and transfusion errors.[4] Among those errors, surgical errors are the most damaging and the costliest. They have harmful effects on patients, including temporary or permanent damage and even significant mortality rates.[5] This affects not only the patients but also their families, burdening them financially. Surgical errors can happen during preoperative, operative, and postoperative processes. Like other complex surgeries, heart surgeries are considered to pose a higher risk of surgical errors due to the intricacies of the heart's anatomical structures and surgical complexities. Reducing surgical errors is necessary to mitigate mortalities and unnecessary costs related to prolonged hospital stays and reoperations. One of the early measurements used to reduce surgical errors in cardiac surgeries was the Institutional Coronary Artery Bypass Surgery Checklist (the InCor-Checklist), which significantly decreased the mortality rate after surgery from 8.2% in 2013 to 3.1% in 2019.[6] Integrating the InCor-Checklist with artificial intelligence will help ensure necessary steps during cardiac surgeries are followed properly. The integration can even decrease the mortality rate and improve the procedure's overall safety and efficiency.

AI and machine learning can be used in heart surgeries during the entire treatment process, from preoperative, real-time surgical monitoring to postoperative predictions. AI assistant copilots can help cardiovascular surgeons plan complex surgeries using 3D models of the heart and surrounding organs to analyze defects and damaged areas. During cardiovascular surgeries, AI can continuously monitor patient data, such as blood pressure and heart rate. If any sudden changes occur, AI can track them and alert the surgical team to adjust treatment, assisting the surgeons in making better and informed decisions. In the postoperative stage, AI helps detect complications through medical imaging before they become severe, leading to shorter recovery time and better patient outcomes.[1] Other applications of AI focus on aiding robotic surgeries in performing complex procedures under surgeons' supervision. This reduces surgical errors and enhances surgical performance. Applications of AI are not limited to only enhancing cardiac surgical accuracies but also avoiding untimely deaths by improving the medication system. According to the European Union study, 1.7 million people aged less than 75

DOI: 10.36838/v7i3.97

This is due to the lack of proper medical knowledge and technology.[7] From these deaths, heart-related diseases take the first place, with more than 180,500 preventable deaths, or 32% of total avoidable deaths. The study indicates that implementing AI and technology in medication can significantly reduce such preventable fatalities by improving early detection, diagnosis, and treatment.

Chan and colleagues analyzed data from the Australian and New Zealand Audit of Surgical Mortality (2009-2015) to identify clinical management issues affecting cardiothoracic surgical care. Nine hundred-eight cases were identified of those 769 that were cardiac-related and determined to have at least one clinical management issue. The analysis included the three most common cardiac surgery procedures: coronary artery bypass grafting, aortic valve replacement, and mitral valve surgery. Of the reported clinical management issues, 47.4% were probably preventable, and 17% were definitely avoidable. These numbers indicate that 64.4% of CMIs were likely preventable if the necessary measurements were taken, particularly for those involving technical issues.[8] The most common preventable CMI for cardiac surgeries was found to be technical issues. These technical issues in intraoperative phases may include but are not limited to, unintentional injury to surrounding anatomical structures, inadequate myocardial protection, and the wrong surgical approach, especially performed by junior trainee surgeons with insufficient supervision.

Technical issues occurring during intraoperative phases and other issues that arise in postoperative phases can be addressed with the augmentation of surgical procedures with AI. By leveraging the capabilities of AI, such as predictive analytics for forecasting potential complications, real-time monitoring, decision support, and anomaly detection, it is possible to prevent intraoperative and postoperative issues. According to Chan *et al*.[8] and The Michigan Society of Thoracic and Cardiovascular Surgeons[9] analysis of mortality in cardiac surgical cases, postoperative phase clinical management issues were the most contributor to mortalities, followed by preoperative and intraoperative issues. It is possible to prevent these issues with improved clinical management practices and appropriate interventions. Postoperative issues such as communication issues and inadequate patient monitoring can be addressed through advanced training focusing on interdisciplinary collaboration, which can improve communication and teamwork skills in surgical settings.[10,11]

The integration of large language models (LLMs) in surgical assistants alongside robotic surgery can significantly enhance the accuracy of cardiac procedures by providing real-time evidence-based guidance, personalized surgical plans, intraoperative monitoring, and predictive analytics for complications during surgery. LLMs' capability to assist surgeons by offering them precise recommendations, flagging potential complications, and optimizing surgical procedures can ultimately streamline workflows and improve patient safety throughout the surgical process. With a combination of surgical robotic systems, LLMs enhance precision and support minimally invasive approaches, leading to less tissue trauma, faster recovery times, and a reduced risk of infections for cardiac patients. Overall, the synergy of AI-based surgical assistants and robotics holds immense potential in improving both the safety and efficiency of surgical procedures, leading to better long-term outcomes.

While numerous studies have explored the impact of robotic-assisted cardiac surgeries on enhancing surgical outcomes, the integration of LLMs in cardiac surgeries is in its nascent stage and remains unexplored. This work demonstrates that integrating robotic systems and LLMs with cardiac surgeries can help reduce surgical errors and improve surgical outcomes. We first define artificial intelligence and its applications in surgical procedures. We then discuss robotic-assisted surgery, how it works, its historical background, and the benefits it brings to surgeons and patients. Then, we compare mortality rates, cost, duration of surgery, intensive care unit (ICU) stay, length of hospital stay, and complications for robotic-assisted and nonrobotic cardiac surgeries. Subsequently, we introduce two large language model-based systems (1) ChatGPT and (2) CARES Copilot 1.0, and discuss how they can be used in surgical procedures during preoperative, intraoperative, and postoperative phases of surgeries. Additionally, we examine the challenges and limitations of the systems preventing their widespread adoption in surgical procedures. Lastly, we outline our conclusion and discuss future research.

Artificial Intelligence in Cardiac Surgery:

Artificial intelligence (AI) is a system that models human reasoning to solve a wide array of problems, from simple tasks such as recognizing text and speech to complex tasks such as large-scale data analysis and performing intricate surgeries.[12] AI systems extensively use large datasets in their training process to make predictions and draw conclusions based on patterns in the sample training data. In today's world, AI is integrated into our daily activities, helping users complete tasks more efficiently. In the medical industry, AI systems are improving the efficiency and efficacy of clinical outcomes. Within surgery, AI and machine learning algorithms are employed throughout the surgical process, from planning operations to predicting postoperative complications.[13] Additionally, AI algorithms help develop personalized surgical plans by considering individual patient data, such as medical history and genetic information. This benefit enables surgeons to tailor surgical plans for each patient, lowering the risk of complications.

1. Robotic-assisted cardiac surgery:

Robotic surgery is often considered as a robot performing surgical procedures autonomously without human interaction; however, the robots are guided by surgeons to perform complex surgeries with high precision. Robotic surgery is a minimally invasive surgery that uses robotic systems, which consist of a console where the surgeon sits and controls the robotic arms, equipping tiny surgical tools through the joysticks that translate the surgeon's hand movements into highly accurate actions by the robotic arms.[14] The robotic arms can make precise movements with tiny incisions, reducing pain, less tissue damage, less blood loss, less scarring, faster recovery times, and fewer risks. The system uses a high-definition camera, providing a magnified and enhanced view of the surgical site. The Da Vinci® Surgical System by Intuitive Surgical, Ion by Intuitive Surgical, and Mako by Stryker are the most common surgical robotic systems.[15] Several procedures use robotic surgical systems, including cardiologic, colorectal, urologic, gynecologic, thoracic, head and neck (transoral), metabolic, and intra-abdominal surgery.[14] Numerous studies have indicated that robotic-assisted cardiac surgeries offer significant benefits, including reduced hospital stays and fewer complications than non-robotic surgeries.[16]

In 1995, Intuitive Surgical manufactured the da Vinci® surgical system (Figure 1) in California, which uses a minimally invasive approach for complex surgeries. Approved by the Food and Drug Administration (FDA), the da Vinci® surgical system was initially developed for cardiac surgeries but has since been deployed in various types of surgeries, including urologic, gynecologic, pediatric, colorectal, and general surgeries.[17] On May 7, 1998, using the da Vinci® surgical system, Carpentier and Loulmet performed the first open-heart robotic cardiac surgery, an endoscopic aortocoronary bypass, in Paris to repair an atrial septal defect via mini-thoracotomy.[18] Because of the success of the surgery, it opened a new era of using robots in open-heart cardiac surgeries. More than 750,000 surgeries have been performed using the da Vinci® surgical system[19] since its approval from the FDA in 2000, enhancing surgical procedures and producing better patient outcomes. It has been more than 25 years since the first robots entered the surgical world, but their adoption in cardiac surgery has been slow. This situation was due to skepticism and resistance to change within the surgical community since there were failed early operations and the high costs of surgical robotic systems and their maintenance costs.[20] Heart transplantation, transcatheter aortic valve replacement (TAVR), aortic stenosis, coronary artery bypass graft (CABG), postoperative atrial fibrillation (POAF), and robotic mitral or tricuspid valve surgery are the most common cardiac surgeries where robotic procedures are utilized.[13,21]

Figure 1: The da Vinci® surgical system

In comparisons of outcomes of robotic-assisted procedures versus traditional cardiac approaches, the observed differences are significant, especially for mitral valve surgeries and coronary artery bypass grafting (Table 1). One of the few consistent findings from the various extant studies has been the fact that robotic-assisted procedures register much lower rates of all-cause mortality and fewer complications of any kind compared to non-robotic surgery. Chi *et al.*[16] and Coyan *et al.*[22] found that the mitral valve procedures with the use of robotics resulted in a 0% mortality, while with conventional, non-robotic surgeries, the mortality rates were much higher. Hawkins *et al.*[23] and Wang *et al.*[24] also reported that robotic surgery, as compared to conventional surgeries, reduces the duration of both ICU and hospital stays. This means that robotic interventions are associated with better chances of recovery. Although these robotic procedures could consume more operative time (Hawkins *et al.*,),[23] total patient outcomes, especially in recovery and complications associated with each procedure, depict a considerable advancement.

While robotic surgery is associated with significant benefits, it is not without its limitations. Importantly, numerous studies have revealed that the length of hospital stays and ICU stays don't reduce significantly compared to conventional methods (Chemtob *et al.*,; Yanagawa *et al.*,).[25,26] While fewer complications are evident following robotic surgery than conventional procedures, according to studies conducted by Yanagawa *et al.*[26] and Hawkins *et al.*,[23] complications can arise. These mixed outcomes can be attributed to the precision of the robotic equipment and the learning curve associated with implementing and handling robotic systems. With that, even though robotic-assisted cardiac operations have proved to be effective in many ways—such as reduction in mortality and improvement in most aspects of recovery—there are still elements in this technology that require improvement to eliminate variations in surgical efficiency and reduce problems.

Table 1: This table summarizes all-cause mortality rates, complication rates, intensive care unit (ICU) stays (mean hours), and length of hospital stays (LOS, mean days) of robotic-assisted and conventional (non-robotic) cardiac surgeries, including coronary artery bypass grafting (CABG), mitral valve repair/replacement surgery (MVS), and general cardiac surgeries.

No	Study	Year	Procedure	n (R)	n (NR)	All-Cause Mortality (%): (R)	All-Cause Mortality (%): (NR)	Complication (%): (R)	Complication (%): (NR)	ICU Stay (hours): (R)	ICU Stay (hours): (NR)	LOS (days): (R)	LOS (days): (NR)
1.	Chemtob et al.[25]	2022	MVS	605	395	0	0	Lower	Higher	30.3	34.0	5.2	5.9
2.	Hawkins et al.[23]	2018	MVS	372	1,352	0.8	3.4	9	9	31.7	57.7	4.7	6.7
3.	Sicim et al.[27]	2021	MVS	64	66	3.1	3.0	1.6	3.0	38.4	62.4	7.9	9.4
4.	Yokoyama et al.[28]	2021	CABG	7,355	1,204,125	1.1	2.1	Lower	Higher	-	-	7.3	9.3
5.	Leyvi et al.[29]	2014	CABG	150	1,619	0	1.8	2.7	11.5	-	-	6.0	9.0
6.	Yanagawa et al.[26]	2015	General Cardiac Surgeries	5,199	10,331	Aortic Valve: 0	Aortic Valve: 1.7	27.2	30.3	-	-	Aortic Valve: 4	Aortic Valve: 6
						Mitral Valve: 1.2	Mitral Valve: 2.1					Mitral Valve: 5	Mitral Valve: 6
						Cardiac Vessel: 1.0	Cardiac Vessel: 1.7					Cardiac Vessel: 5	Cardiac Vessel: 6
7.	Chi et al.[16]	2022	MVS	23	44	0	9.0	-	-	3.0	5.0	12.0	17.0
8.	Wang et al.[24]	2018	MVS	503	503	0.8	2.4	35.1	52.9	42.0	61.8	5.3	6.9
9.	Dokollari et al.[30]	2024	CABG	267	267	2.6	5.2	Lower	Higher	31.2	65.0	5.0	7.0
10.	Coyan et al.[22]	2018	MVS	91	91	0	2.2	49.5	53.8	27.5	34.0	5.0	7.0

The costs of robotic-assisted and nonrobotic cardiac surgical procedures were compared across different surgical procedures (Table 2), including coronary artery bypass grafting (CABG), mitral valve surgery (MVS), aortic valve surgery, and cardiac vessel surgery. Yokoyama *et al.*)[27] and Dokollari *et al.*[28] compared the costs of robotic-assisted and non-robotic CABG surgeries. The costs of robotic cardiac surgical procedures were significantly lower than non-robotic surgeries with an approximate decrease in costs of $4,879 for Yokoyama and $16,854 for Dokallari respectively (Yokoyama and Dokollari). Particularly, Dokollari and colleagues' recent data indicated definite

DOI: 10.36838/v7i3.97

financial benefits from robotic-assisted CABG surgeries. In some studies, the decreased cost associated with robotic procedures is due to fewer complications, transfusion requirements, readmission rates, hospital ICU stays, and hospital stays.[16,29] For example, Alemozaffar *et al*.[29] compared hospital costs associated with laparoscopic and robot-assisted partial nephrectomy. The robotic-assisted surgery resulted in shorter operating room time and length of hospital stays, decreasing the mean costs with better patient outcomes.[29] However, research by Coyan *et al*.,[22] Paul *et al*.,[30] and Suri *et al*.[31] demonstrates that total costs between robotic and conventional MVS are less pronounced and comparable. In contrast, Yanagawa *et al*.[26] reported higher costs for robotic surgeries across several cardiac procedures including aortic valve, mitral valve, and cardiac vessel surgeries. The differences were as high as $4,280 for the mitral valve, $4,943 for the cardiac vessel, and $3,788 for the aortic valve surgeries, which were significantly higher as compared to nonrobotic procedures. This suggests that the cost-effectiveness of robotic-assisted surgeries may vary based on the specific procedure either mitral, CABG, aortic, or overall procedures depending on their cost, with some of the cases favoring the non-robotic methods. Overall, data suggests that while robotic-assisted surgeries can be more cost-effective in specific cardiac surgery types like CABG, this is not necessarily true for all cardiac procedures.

Drawing meaningful conclusions from the cost comparison of robotic to non-robotic surgery is challenging for multiple reasons. It is difficult to quantitatively evaluate whether the decrease in complications and shorter ICU and hospital stays offset the costs of robotic-assisted cardiac surgeries.[29] Most of the studies did not consider the initial capital investment of robotic equipment, maintenance, implementation period, and direct technical hospital costs. Additionally, factors such as the impact of improved ergonomics on surgeon longevity and shorter intervals between surgery and the patient's return to productive employment should be considered in cost comparisons. In the previous section of the paper, we reviewed the role of robotic-assisted surgery in reducing cardiac surgical errors. In the next section, we will discuss LLM-based surgical assistants and how they can be integrated into surgical procedures.

Table 2: This table summarizes the costs of robotic-assisted and conventional (non-robotic) cardiac surgeries, including coronary artery bypass grafting (CABG), mitral valve repair/replacement surgery (MVS), and general cardiac surgeries. Costs are presented in US dollars.

	Study	Year	n(R)	n(NR)	Procedure	Robotic (Cost, $)	Non-Robotic (Cost, $)	p =
1.	Yokoyama et al.[27]	2021	7355	1,204,125	CABG	41,317	46,196	<0.01
2.	Dokollari et al.[28]	2023	267	267	CABG	18,726	35,580	
3.	Coyan et al.[22]	2018	91	91	MVS	27,662	28,241	0.27
4.	Paul et al.[30]	2015	3,145	47,263	MVS	33,720	34,509	0.44
5.	Suri et al.[31]	2013	185	185	MVS	32,144	31,838	0.32
6.	Yanagawa et al.[26]	2015	5199	10331	Aortic Valve	42,366	38,578	0.006
					Mitral Valve	43,099	38,819	<0.001
					Cardiac Vessel	43,521	38,578	<0.001

Notes. CABG = coronary artery bypass grafting; MVS = mitral valve surgery.

2. Large language model-based systems in cardiac surgeries:

Beginning from the introduction of machine learning concepts by Alan Turing in 1950, AI has dramatically evolved over the past seven decades. The invention of transformers like Bidirectional Encoder Representations from Transformers (BERT) and self-attention culminated in the development of natural language processing. Using these transformers, especially BERT, was pivotal in the advancement of LLMs that generate coherent and well-structured texts.[32] Large language model evolution has now reached the era of GPT-4, Llama, and other models with multibillion parameters, making them capable of processing large amounts of reasoning tasks with higher computational capabilities.

Large language models are computational models that can understand and generate human-like language based on input by predicting the likelihood of word sequences.[33] They are trained on in-context learning to generate text based on a given context or prompt, making them generate coherent and contextually relevant responses with great accuracy.[34] LLMs can learn from mistakes using reinforcement learning from human feedback—a machine learning technique involving fine-tuning the model using human-generated responses as rewards based on their responses.[28] This improves the model's performance in generating accurate text. GPT-4, Claude, BERT, Llama 3.1, and Gemini are examples of language models. LLMs have a wide range of applications, including but not limited to natural language processing tasks like content summarization, translation, reasoning, sentiment analysis, and question answering, as well as synthesizing large amounts of data, content creation, and automating services. Furthermore, the impact of LLMs extends to diverse fields, such as law, medicine, robotics, social sciences, computational biology, and so on.[28] In the following sections, we will explore the applications of two LLM-based surgical assistants in cardiac surgeries and how they help mitigate surgical errors.

A. Applications of ChatGPT in Cardiac Surgery:

ChatGPT is a large language model launched by OpenAI in November 2022 that uses GPT (generative pretrained transformer model) technology, which plays significant roles in several industries. GPT models are developed to generate human-like responses for tasks such as summarizing texts and answering questions.[35] The medical industry uses ChatGPT to support medical education, summarize patient diagnosis reports, keep healthcare providers updated with the latest research, and assist with telemedicine.[36] ChatGPT evidently has the potential to enhance the medication process and assist cardiac surgeons.

ChatGPT has the following potential applications in heart surgery:

1. Preoperative planning and surgical simulation: Heart surgery is complex and requires meticulous planning, preparation, and risk assessment. ChatGPT can help with preoperative planning by summarizing patients' medical histories, genetic histories, and electronic health records to provide the surgical team comprehensive understanding of their conditions and potential risk factors. This helps decrease the time that surgeons commit to a single patient and allows them to see more patients who are requesting their services than they otherwise could. ChatGPT can also be used in surgical simulations to train surgeons, visualize surgical scenarios in advance, and predict potential complications.[37]

2. Intraoperative monitoring: ChatGPT can monitor patient data in real-time, including heartbeat, blood pressure, anesthesia levels, and blood loss. This provides surgeons with opportunities to spot anomalies in the surgery and readjust operation procedures. Another surgical error that ChatGPT can help with is administering the correct anesthesia dose by analyzing empirical patient data, such as age, weight, allergies, and anesthesia records. This helps anesthetists respond efficiently to the risk of complications such as overdosing, underdosing, or adverse reactions.[37,38]

3. Decision-making support: Instantaneous decision-making is critical during surgery. Surgical teams can use ChatGPT to make informed decisions founded on recent research that is less resilient on personal experiences. This can reduce surgeons' cognitive biases that may be caused by a lack of necessary information such as patient medical history, surgical risk factors, and best practices and guidelines, which can influence surgery outcomes.[39,40]

4. Education and training: Large language models, such as ChatGPT, are crucial learning tools for students. ChatGPT is trained on many textbooks, literature, and internet data, so it can answer questions for surgeons with high accuracy, depending on the most recent information available to it. Trainees can access extant surgical case studies and step-by-step procedural guides using ChatGPT's platform. In medical research, ChatGPT can be used to analyze medical records to predict patient outcomes and support personalized medicine research via its ability to help researchers develop medications that are tailored to an individual's specific circumstance, leading to better patient outcomes.[41]

5. Postoperative care: Postoperative care is crucial in helping patients recover completely.[38,42] Insufficient or poorly managed postoperative care can result in complications for patients. ChatGPT can help doctors develop comprehensive, personalized rehabilitation plans for postoperative patients and choose appropriate medications. ChatGPT's postoperative roles include scheduling doctors' appointments, identifying patient symptoms and recommending treatments or therapies, medication reminders with the appropriate doses and possible side effects, providing instant advice to patients as needed, scheduling virtual follow-up consultations, and helping doctors with the documentation and reporting processes.[42]

To use ChatGPT in surgical practices, it is necessary to fine-tune the model with domain-specific surgical data, expert feedback from cardiologists, and up-to-date medical guidelines and protocols. Fine-tuning means training pre-existing models on a specific dataset to enhance their performance in a particular domain. Instead of training a new model for a specific task, fine-tuning provides a cost-effective approach to leverage pre-acquired knowledge of base models, like GPT and Llama.[43] Fine-tuning ChatGPT for surgical procedures can provide surgeons with more accurate information on surgical techniques, best practices, and guidelines while minimizing errors and inaccurate responses, leading to improved patient care. However, the cost of fine-tuning became a significant challenge to its adoption. Additionally, biases on ChatGPT related to surgeries are a big concern, leading to misdiagnosis, disparities in treatment recommendations, and potential harm to patient safety.[37] Another significant challenge of ChatGPT and other LLMs is hallucinations—inaccurate, misleading, fabricated, or contextually irrelevant information generated by large language models. Hallucinations occur due to the model's reliance on the patterns learned during the training rather than access to real-time and verified information. This can lead to substantial complications and risks, especially in medicine, if not addressed effectively. Furthermore, the inability to understand patients' sentiments' is another limitation of AI models. As they only consider patients' quantitative data, their emotions, and psychological status are not considered, negatively impacting the overall treatment process.[44] In this section, we discussed the potential applications of ChatGPT in heart surgeries. In the next section, we will discuss CARES Copilot 1.0, a large language model specifically developed to assist surgeries.

B. CARES Copilot 1.0: The Game-Changing AI Surgical Assistant:

CARES Copilot, officially launched globally at the Huawei CONNECT conference 2023, aims to improve the safety and efficiency of surgical operations. CARES Copilot 1.0 (Figure 2) is an advanced large language model-based surgical assistant developed by the Centre for Artificial Intelligence and Robotics (CAIR). Based on Meta Platforms Inc.'s Llama 2.0 large language model, CARES was trained and fine-tuned on over 3,000 pages of literature and optimized by surgical professionals.[45-47] It outperformed medical exams, achieving state-of-the-art performance. According to the CARES Copilot 1.0 research team, the copilot was rigorously tested using a challenging zero-shot top token selection standard (output selection strategy) across five public datasets and two proprietary neurological knowledge evaluation datasets.[48] With the ability to handle more than 100,000 tokens, CARES Copilot can rapidly understand surgical textbooks and medical documents within seconds and generate accurate answers.[49,50] Its integration with medical devices and capability to process multimodal surgical data—including images, text, voice, video, MRI, ultrasound, and CT scans—makes it an ideal solution for complex surgeries like heart surgeries, boasting an impressive 95% accuracy rate as tested on brain surgeries in more than seven hospitals across Hong Kong and the Greater Bay area.[45]

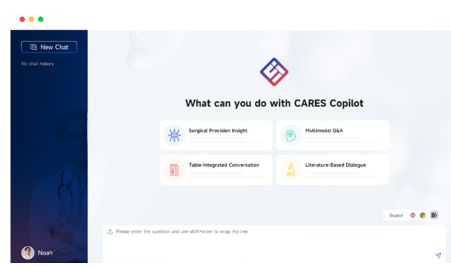

Figure 2: CARES Copilot 1.0 Chatbot. Extracted from "CARES Copilot,"[49] https://www.cares-copilot.com/ (accessed August 24, 2024).

DOI: 10.36838/v7i3.97

CARES is considered the world's first real-time simulator for neurosurgery. With real-time brain-instrument interaction, CARES can simulate brain surgeries to enhance surgical precision, reduce errors, provide a risk-free training process for surgeons, and increase the confidence of the surgical team while reducing stress.[49] Another significant advantage of CARES is its lightweight private deployment capability, which facilitates its use in hospitals and surgical rooms. Its easy deployment makes CARES a potential solution for low- and middle-income countries, where many patients are waiting for enhanced and safe cardiac surgical care.

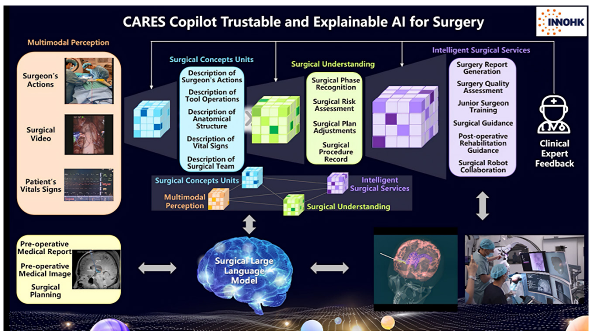

Figure 3: CARES Copilot 1.0: A Trustworthy and Explainable AI Framework for Enhancing Surgical Decision-Making and Outcomes. Extracted from "Multimodal AI Large Model,"[51] https://www.cair-cas.org.hk/articles/ai (accessed August 24, 2024).

At the preoperative stage, CARES uses various cameras to monitor medical staff and ensure all pre-operation preparations are complete. (Figure 3 shows the overall CARES Copilot framework, while Figure 4 illustrates its integration into surgical procedures.) During operations, CARES uses endoscopy video to track surgical progress and simulate tool and tissue segmentation.[47] The system leverages computer vision and deep learning techniques to support decision-making during surgeries. By integrating endoscopy and AI, CARES tracks every step of the surgical procedure and provides surgeons with detailed real-time feedback and operational reminders, warning surgeons from taking risky procedures to enhance precision and reduce the likelihood of errors. Despite its capabilities, researchers want CARES Copilot to take more active roles during surgeries and develop more reliable ones that doctors can rely on. However, the lack of access to the cutting-edge Nvidia's advanced chips restricted its computing power and expansion to more advanced surgeries, hindering its development.[46]

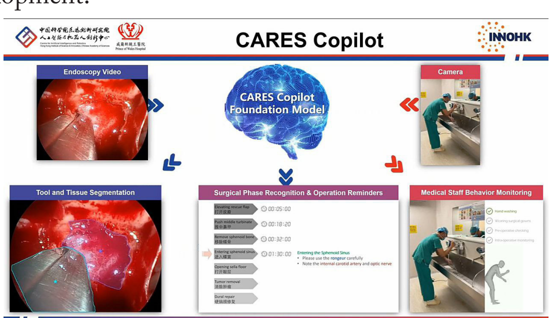

Figure 4: Overview of CARES Copilot Foundation Model Integration in Surgical Procedures. Extracted from "Multimodal AI Large Model,"[51] https://www.cair-cas.org.hk/articles/ai (accessed August 24, 2024).

Limitations of AI-Enabled Systems in Cardiac Surgeries:

Artificial intelligence systems in healthcare usually use automated decision-making techniques, including machine learning, natural language processing, and convolutional neural networks to perform tasks in the clinical setting, which usually require human intelligence. These systems also use huge amounts of data to provide the best outcomes, making them highly reliant on the model training data. Based on these techniques, AI chatbots and surgical robots are continuously helping surgeons make well-informed decisions during surgical procedures while minimizing significant amounts of time and clinical errors. AI assistants are also potential solutions for places where there is a scarcity of resources and educated experts in the surgical field. Their capability of saving the time and energy to perform experiments requires huge amounts of research literature that is probably beyond individual expertise.[52]

Although its utility is apparent, the application of AI in medicine has been the point of contention among many scientists.[53] Some justify using AI to support efficient and effective diagnosis and guide appropriate treatment. The proponents of the use of AI in medicine endorse the use of AI based on its ability to store vast amounts of data that can subsequently be analyzed to suggest proper diagnosis and treatment. Other researchers and medical doctors do not agree with the current application of AI in medicine. Some opponents are concerned that, in the future, AI is going to take over from a medical practitioner by reducing his or her potential to practice medicine using a humanistic touch.[53] Instead of adopting AI to create diagnosis and treatment, opponents argue that AI is intended to supplement medical decision-making where the final decisions are made by the human practitioner. But for others, AI is perceived to be a threat to medicine and the treatment of patients. Some of the limitations of AI include hallucinations, biases, job displacement, and ethical concerns, preventing their wider application in the surgical setting. In this section of the paper, we will discuss the two most significant limitations of AI-enabled systems: hallucinations and biases in cardiology and healthcare as a whole.

1. Hallucinations

What is an AI hallucination?:

AI hallucination, as defined by ChatGPT-4 (August 7, 2024),

"An AI hallucination occurs when an artificial intelligence model generates information or statements that are plausible-sounding but factually incorrect, misleading, or entirely fabricated. These errors can appear in various forms, such as false facts, imaginary sources, or nonsensical responses that seem coherent at first glance. AI hallucinations are more common in complex tasks, like language generation, where the model "guesses" rather than retrieves or calculates the exact answer. These hallucinations typically arise because AI models generate responses based on patterns in data rather than true understanding. The AI's algorithms prioritize producing contextually relevant and syntactically correct responses, which can sometimes result in convincing but erroneous outputs."

DOI: 10.36838/v7i3.97

One of the biggest concerns with modern large language models is their tendency to produce hallucinations. AI hallucinations are instances where erroneous, delusional, or fabricated responses that are not based on any information or data on which the models are trained.[54] Recently, ChatGPT has hallucinated by generating fake legal opinions, which led a law firm to get into trouble for submitting these fabricated claims.[55,56]

In 2023, Aljamaan *et al.*[32] challenged medical AI chatbots with medical prompts (inputs) and reported a reference hallucination score (Table 3), aiming to evaluate the authenticity of AI chatbots' citations. Hallucinations were observed in all parts of the references, including article title, journal name, author name, DOI (Digital Object Identifier), and reference web link. These results imply that chatbots are likely to generate irrelevant, inaccurate, or nonexistent references when asked about complex medical topics. Among the chatbots examined, ChatGPT (GPT-3.5) and Bing were reported to have the highest reference hallucination score of 11, which reflects worse performance compared to other chatbots, such as SciSpace and Elicit, which have a score of 1. ChatGPT's performance in reference generation raises concerns about its application, particularly in education, the training of surgeons, and medical research.

Table 3: Reference Hallucination Scores with Bivariate Analysis of the Artificial Intelligence Chatbots. Adapted from Aljamaan *et al.*[32]

Reference identifier	ChatGPT	Perplexity	SciSpace	Elicit	Bing	Chi-square *(df)*	*P* value
Title	82	41	0	0	46	231.4 (4)	<.001
Digit Object Identifier	89	73	8	4	53	261.26 (4)	<.001
Journal name	83	56	0	0	49	249.4 (4)	<.001
Authors' names	89	74	0	0	65	341.1 (4)	<.001
Publication date	88	48	0	40	61	199.9 (4)	<.001
Reference web link	84	49	3	1	50	231.4 (4)	<.001
Reference relevance to topic prompt	77	63	60	58	50	10.77 (4)	.03

AI chatbots produce varying degrees of hallucinations in content generation, as shown in Figure 5. Chatbots with higher degrees of hallucinations cannot be reliably deployed in the healthcare domain, where a higher degree of accuracy is paramount. As indicated in the LLM hallucination leaderboard below, LLMs like GPT-4 (ChatGPT) and Llama 2 (CARES Copilot 1.0) have fewer hallucinations and higher accuracy rates, making them preferable for surgical and clinical applications. Though ChatGPT performed worse in reference generation, it has the lowest overall hallucination rate in various fields, according to HHEM 2.1, a hallucination detection model (Figure 5).

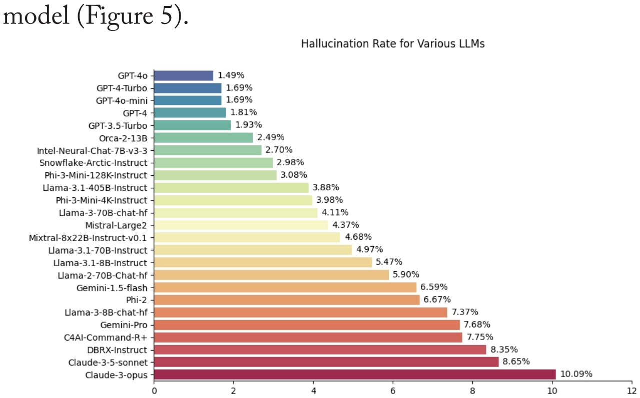

Figure 5: HHEM-2.1 Hallucination Evaluation Model LLMs Hallucination Leaderboard. Extracted from "HHEM 2.1: A Better Hallucination Detection Model and a New Leaderboard,"[57] https://www.vectara.com/blog/hhem-2-1-a-better-hallucination-detection-model (accessed August 7, 2024).

Various methods can mitigate hallucinations. Prompting techniques like chain-of-thought, tree-of-thought, and self-consistency have enhanced LLM reasoning ability by using logical steps to answer each prompt and cross-check its answers.[58] Retrieval-augmented generation architecture, which uses relevant documents from external sources such as search engines, databases, and news feeds, helps LLMs generate more accurate, up-to-date, and verifiable answers. Fine-tuning is another technique commonly employed in LLMs to mitigate hallucinations in domain-specific contexts.[59,60] It uses high-quality and reliable datasets to adapt general-purpose LLMs, like GPT-4 and Llama, toward specific domain models,[61,62] including CARES Copilot 1.0—which is based on Llama. These techniques and other strategies will help develop coherent and reliable models for surgical applications.

Xu *et al.*[63] and other researchers have concluded that it is impossible to completely avoid hallucinations from LLMs, as they are inevitable and innate characteristics. No matter how LLMs are trained, they always miss some possibilities, just as we cannot list out all real numbers (infinite set) even if we try.[63] When LLMs face novel and complex prompts (inputs), they can make errors by retrieving data that seems right but is not actually true. This is because they are highly reliant on the data patterns on which they are trained. Additionally, Xu *et al.* claimed that restricting LLMs to simpler models will still result in errors at certain complex inputs, as their computational ability is limited to the information they can process and verify. Even if it is impossible to eliminate hallucinations with LLMs, it is crucial to use models with fewer hallucinations to widely deploy artificial intelligence in the cardiac surgical field.

2. Biases:

In the era of artificial intelligence, where most clinical practices are aided by AI systems, AI biases widen and exaggerate existing healthcare disparities, benefiting only certain groups with their highest performance.[64] Historically, women and minority groups in the US—Black and African Americans, Hispanics, Asian Americans, and Native Americans—are underrepresented in cardiology in terms of research studies, treatment, and clinical outcomes.[65] AI models are trained with real data from previous medical experiences and trends, so they tend to reflect biases that were once part of clinical practices. Bias in AI usually manifests as the occurrence of unequal or disproportionate outcomes of AI systems. These disparities usually result from factors such as race, gender, and socioeconomic status. For example, an AI model used for risk prediction for heart surgeries may not perform well for Black and underrepresented patients if the model is primarily trained using data from predominantly White populations. AI biases widen and exaggerate existing healthcare disparities, causing only certain groups to benefit from AI's highest performance.

Strategies to Mitigate Biases:

Biases may occur at each level of the model's development steps, including data collection and preparation, model training, validation, and deployment. Therefore, bias-mitigating

DOI: 10.36838/v7i3.97

strategies must focus on these developmental stages and may include, but are not limited to, the following approaches:

Using Heterogeneous Datasets: It is important to ensure that AI models are trained on diverse and representative datasets, including a wide range of geographic and socioeconomic groups.[66] Using data sampling techniques like oversampling underrepresented groups or undersampling overrepresented groups helps to balance the dataset before training.[67,68] Employing these data sampling methods helps reduce class imbalance (population) by changing the dataset distribution through repeating/dropping samples, respectively. In 2021, the FDA published the Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan[69] and the Good Machine Learning Practice for Medical Device Development: Guiding Principles (GMLP).[66] The FDA's guidelines recognized the challenges and importance of addressing biases when developing AI or ML medical devices and systems.

Model Training: During model training, fairness-conscious algorithms such as adversarial debiasing or fairness constraints[70] can adjust for disparities in model performance across different demographic groups. Certain clinical features such as race, gender, or socioeconomic status affect the performance of AI models if not used appropriately. Using machine learning methods such as adversarial learning, which works by training models to minimize the performance of a "discriminator" tasked with identifying protected attributes (such as race or gender), helps to ensure that the model does not rely on these attributes to make predictions.[71] Prioritizing other clinically relevant features (e.g., comorbidities, laboratory test results, and physiological measurements) overprotected features helps mitigate bias, in the model training phase.

Deployment and Monitoring: During the post-development phase, developers continuously monitor and assess AI systems performance, which helps them identify any emerging disparities in the real clinical setting.[66] After the deployment of AI-enabled systems, tracking patients' outcomes by demographic groups in the system helps pinpoint changes in patient population dynamics. Evaluating model performance not only across the overall population, but also across specific populations—women, racial minorities, and elderly patients—ensures equitable assessment on how well the model generalizes across each demographics.[72]

Discussion

This paper highlights critical applications of AI-based systems for preventing surgical errors in cardiac surgical care by introducing ChatGPT and the novel surgical assistant—CARES Copilot 1.0. The potential applications of AI in cardiac surgery range from preoperative planning and risk assessment to intraoperative real-time surgical support and postoperative rehabilitation care. With ongoing research and advancement in AI, ChatGPT, and CARES Copilot 1.0 have immense potential to transform surgical care, enhance surgical training, and support medical research. As CARES Copilot 1.0 has been successfully implemented in neurosurgeries, demonstrating its ability to optimize surgical procedures with higher efficiency, it should also be used in cardiac surgeries to reduce complications and yield better patient outcomes. By augmenting LLM-based surgical assistants (such as CARES Copilot and ChatGPT) with robotic surgical systems, it is possible to reduce the risks of surgical errors throughout the procedure and enhance overall patient outcomes.

Robotic technology is rapidly advancing in the surgical field. Autonomous and semi-autonomous robotic systems are increasingly investigated in surgical procedures, from low-level automation early surgical robots to high-level autonomous robots performing complex endoscopic operations. For example, in 2022, the Smart Tissue Autonomous Robot (STAR)—developed by Johns Hopkins University—performed the first fully autonomous laparoscopic surgery in four experiments in pig tissues.[73] STAR performed intricate surgeries with higher accuracy and consistency than humans performing the same procedures. It is expected that soon AI-driven autonomous robots will enter the cardiac surgical field and improve the outcomes of surgeries.[12] By reducing the costs of robotic surgical systems, it is possible to replace conventional cardiac surgeries with robotic ones while reducing mortalities, surgical errors, and unnecessary costs. Future research will require an interdisciplinary effort involving biomedical engineers, AI developers, and medical professionals to develop highly reliable surgical systems. However, AI-based systems should not replace healthcare professionals but augment them in providing better patient care. Therefore, while leveraging the benefits of AI in cardiac surgery, it is crucial to adopt the human-AI approach[66] in surgical procedures where the central decision-making remains to surgical professionals. This approach will help us medical professionals' skills and expertise, safeguard their roles, and preserve jobs in the medical field.

Future work should focus on new ways of reducing biases, particularly in surgery, that can lead the model to amplify biases from training datasets, reflecting historical disparities and inequities in healthcare care. These biases may include cultural, gender, or socio-economic biases, impacting surgical care by providing misdiagnosis and unequal treatment recommendations.[37] Additionally, the medical AI field should aim to develop AI systems with reduced hallucinations, reducing costs related to fine-tuning LLMs, and producing more surgical systems to address accessibility of cardiac surgical care worldwide. This way it is possible to create systems that are efficient and highly reliable. Moreover, government officials and policy-makers should set clear policies and regulations regarding job displacement resulting from artificial intelligence. As AI advances, it is crucial to balance its significant benefits with workforce protection to ensure that AI systems augment with humans, rather than replace their expertise.

Conclusion

Overall, robotic-assisted cardiac surgery has significantly reduced complications, mortality rates, and length of hospital stays compared with conventional surgeries. Furthermore, the wider integration of robotic surgery and AI systems will bring a transformation in surgical precision and patient outcomes. AI's real-time surgery simulation will help track every step involved and prevent surgical errors through operation remainders and real-time surgery feedback. AI-based surgical ssistants, such as CARES Copilot 1.0 and ChatGPT, can potentially integrate with cardiac surgery and provide state-of-the-art surgical support to the surgical team in the entire surgical process, ensuring the accuracy and efficiency of surgical operations. Further research will help improve LLMs' capability in surgical procedures and address issues related to hallucinations, biases, and deployment costs to make heart surgeries safe and easily accessible for all.

Acknowledgments

First and foremost, I would like to thank God (ልዑል እግዚአብሔር፡ ምስጋና ይገባሃል። – “Almighty God, you deserve praise”) for his guidance and blessings throughout this journey. Special thanks to my mentor, Professor Siddharth Krishnan (Department of Computer Science at the University of North Carolina—Charlotte), for guiding me in the writing process in all phases of my research project and for his unwavering support in understanding machine learning, large language models, and AI-based systems. I also wish to acknowledge Billy Gao (Stanford University) for mentoring and assisting in formatting and editing my paper. Additionally, I want to thank both my parents (Getenet Bekele and Abebaye Alemayehu), as well as my grandmother, Gorfe Tessema, for their unwavering support and assistance throughout all stages of my research and academic pursuits.

References

1. Alhiti, H. A. R. Artificial Intelligence Enhances Heart Surgeries. *South East Eur. J. Cardiol.* **2023**, *4* (1), 67–69. https://doi.org/10.3889/seejca.2023.6056.
2. Vervoort, D.; Lee, G.; Lin, Y.; Contreras Reyes, J. R.; Kanyepi, K.; Tapaua, N. 6 Billion People Have No Access to Safe, Timely, and Affordable Cardiac Surgical Care. *JACC Adv.* **2022**, *1* (3), 100061. https://doi.org/10.1016/j.jacadv.2022.100061.
3. Argaw, S.; Genetu, A.; Vervoort, D.; Agwar, F. D. The State of Cardiac Surgery in Ethiopia. *JTCVS Open* **2023**, *14*, 261–269. https://doi.org/10.1016/j.xjon.2023.03.001.
4. Ahsani-Estahbanati, E.; Doshmangir, L.; Najafi, B.; Akbari Sari, A.; Sergeevich Gordeev, V. Incidence Rate and Financial Burden of Medical Errors and Policy Interventions to Address Them: A Multi-Method Study Protocol. *Health Serv. Outcomes Res. Methodol.* **2022**, *22* (2), 244–252. https://doi.org/10.1007/s10742-021-00261-9.
5. Mehtsun, W. T.; Ibrahim, A. M.; Diener-West, M. Surgical Never Events in the United States. *J. Vasc. Surg.* **2013**, *58* (1), 277. https://doi.org/10.1016/j.jvs.2013.05.040.
6. Mejia, O. A. V.; de Mendonça, F. C. C.; Sampaio, L. A. B. N.; Galas, F. R. B. G.; Pontes, M. F.; Caneo, L. F.; Dallan, L. R. P.; Lisboa, L. A. F.; Ferreira, J. F. M.; Dallan, L. A. de O.; Jatene, F. B. Adherence to the Cardiac Surgery Checklist Decreased Mortality a Teaching Hospital: A Retrospective Cohort Study. *Clinics* **2022**, *77*, 100048. https://doi.org/10.1016/j.clinsp.2022.100048.
7. *Are 570 000 deaths per year in the EU avoidable?* https://ec.europa.eu/eurostat/web/products-eurostat-news/-/DDN-20180629-1 (accessed 2024-08-24).
8. Chan, J. C. Y.; Gupta, A. K.; Stewart, S. K.; McCulloch, G. A. J.: Babidge, W. J.; Worthington, M. G.; Maddern, G. J. Mortality in Australian Cardiothoracic Surgery: Findings From a National Audit. *Ann. Thorac. Surg.* **2020**, *109* (6), 1880–1888. https://doi.org/10.1016/j.athoracsur.2019.09.060.
9. Shannon, F. L.; Fazzalari, F. L.; Theurer, P. F.; Bell, G. F.; Sutcliffe, K. M.; Prager, R. L. A Method to Evaluate Cardiac Surgery Mortality: Phase of Care Mortality Analysis. *Ann. Thorac. Surg.* **2012**, *93* (1), 36–43. https://doi.org/10.1016/j.athoracsur.2011.07.057.
10. Chan, J. C. Y.; Gupta, A. K.; Stewart, S.; Babidge, W.; McCulloch, G.; Worthington, M. G.; Maddern, G. J. "Nobody Told Me": Communication Issues Affecting Australian Cardiothoracic Surgery Patients. *Ann. Thorac. Surg.* **2019**, *108* (6), 1801–1806. https://doi.org/10.1016/j.athoracsur.2019.04.116.
11. Feinman, J. Talk About the Echo: The Importance of Effective Communication in the Cardiac Operating Room. *J. Cardiothorac. Vasc. Anesth.* **2020**, *34* (7), 1990–1992. https://doi.org/10.1053/j.jvca.2020.01.050.
12. DIAS, R. D.; SHAH, J. A.; ZENATI, M. A. Artificial Intelligence in Cardiothoracic Surgery. *Minerva Cardioangiol.* **2020**, *68* (5), 532–538. https://doi.org/10.23736/S0026-4725.20.05235-4.
13. Sulague, R. M.; Beloy, F. J.; Medina, J. R.; Mortalla, E. D.; Cartojano, T. D.; Macapagal, S.; Kpodonu, J. Artificial Intelligence in Cardiac Surgery: A Systematic Review. medRxiv July 5, 2024, p 2023.10.18.23297244. https://doi.org/10.1101/2023.10.18.23297244.
14. Bramhe, S.; Pathak, S. S.; Bramhe, S.; Pathak, S. S. Robotic Surgery: A Narrative Review. *Cureus* **2022**, *14*. https://doi.org/10.7759/cureus.29179.
15. *Top 5 Robotic Surgery Systems | Docwire News*. https://docwirenews.com/post/top-5-robotic-surgery-systems (accessed 2024-08-24).
16. Chi, N.-H.; Fu, H.-Y.; Yu, H.-Y.; Wu, I.-H.; Wang, C.-H.; Chou, N.-K. Comparison of Robotic and Conventional Sternotomy in Redo Mitral Valve Surgery. *J. Formos. Med. Assoc. Taiwan Yi Zhi* **2022**, *121* (1 Pt 2), 395–401. https://doi.org/10.1016/j.jfma.2021.05.023.
17. Pugin, F.; Bucher, P.; Morel, P. History of Robotic Surgery: From AESOP® and ZEUS® to Da Vinci®. *J. Visc. Surg*. **2011**, *148* (5 Suppl), e3-8. https://doi.org/10.1016/j.jviscsurg.2011.04.007.
18. Carpentier, A.; Loulmet, D.; Aupecle, B.; Berrebi, A.; Relland, J. Computer-Assisted Cardiac Surgery. *The Lancet* **1999**, *353* (9150), 379–380. https://doi.org/10.1016/S0140-6736(05)74952-7.
19. *About the da Vinci Surgical System | UC Health*. https://www.uchealth.com/services/robotic-surgery/patient-information/davinci-surgical-system/ (accessed 2024-08-24).
20. Jr, W. R. C. Historical Evolution of Robot-Assisted Cardiac Surgery: A 25-Year Journey. *Ann. Cardiothorac. Surg.* **2022**, *11* (6),56482–56582. https://doi.org/10.21037/acs-2022-rmvs-26.
21. Harky, A.; Chaplin, G.; Chan, J. S. K.; Eriksen, P.; MacCarthy-Ofosu, B.; Theologou, T.; Muir, A. D. The Future of Open Heart Surgery in the Era of Robotic and Minimal Surgical Interventions. *Heart Lung Circ.* **2020**, *29* (1), 49–61. https://doi.org/10.1016/j.hlc.2019.05.170.
22. Coyan, G.; Wei, L. M.; Althouse, A.; Roberts, H. G.; Schauble, D.; Murashita, T.; Cook, C. C.; Rankin, J. S.; Badhwar, V. Robotic Mitral Valve Operations by Experienced Surgeons Are Cost-Neutral and Durable at 1 Year. *J. Thorac. Cardiovasc. Surg*. **2018**, *156* (3), 1040–1047. https://doi.org/10.1016/j.jtcvs.2018.03.147.
23. Hawkins, R. B.; Mehaffey, J. H.; Mullen, M. G.; Nifong, W. L.; Chitwood, W. R.; Katz, M. R.; Quader, M. A.; Kiser, A. C.; Speir, A. M.; Ailawadi, G. A Propensity Matched Analysis of Robotic, Minimally Invasive, and Conventional Mitral Valve Surgery. *Heart* **2018**, *104* (23), 1970–1975. https://doi.org/10.1136/heartjnl-2018-313129.
24. Wang, A.; Brennan, J. M.; Zhang, S.; Jung, S.-H.; Yerokun, B.; Cox, M. L.; Jacobs, J. P.; Badhwar, V.; Suri, R. M.; Thourani, V.; Halkos, M. E.; Gammie, J. S.; Gillinov, A. M.; Smith, P. K.; Glower, D. Robotic Mitral Valve Repair in Older Individuals: An Analysis of The Society of Thoracic Surgeons Database. *Ann. Thorac. Surg.* **2018**, *106* (5), 1388–1393. https://doi.org/10.1016/j.athoracsur.2018.05.074.

DOI: 10.36838/v7i3.97

25. Chemtob, R. A.; Wierup, P.; Mick, S. L.; Javorski, M. J.; Burns, D. J. P.; Blackstone, E. H.; Svensson, L. G.; Gillinov, A. M.; Chem tob, R. A.; Wierup, P.; Mick, S. L.; Lopez, D. C.; Javorski, M. J.; S emple, M. E.; Burns, D. J. P.; Blackstone, E. H.; Svensson, L. G.; Desai, M.; Gillinov, A. M. A Conservative Screening Algorithm to Determine Candidacy for Robotic Mitral Valve Surgery. *J. Tho rac. Cardiovasc. Surg.* **2022**, *164* (4), 1080–1087. https://doi.org/10 .1016/j.jtcvs.2020.12.036.
26. Yanagawa, F.; Perez, M.; Bell, T.; Grim, R.; Martin, J.; Ahuja, V. Critical Outcomes in Nonrobotic vs Robotic-Assisted Cardiac S urgery. J*AMA Surg. 2015,* **150** (8), 771–777. https://doi.org/10.10 01/jamasurg.2015.1098.
27. Yokoyama, Y.; Kuno, T.; Malik, A.; Briasoulis, A. Outcomes of R obotic Coronary Artery Bypass versus Nonrobotic Coronary Arte ry Bypass. *J. Card. Surg.* **2021**, *36* (9), 3187–3192. https://doi.org/ 10.1111/jocs.15710.
28. Dokollari, A.; Sicouri, S.; Prendergrast, G.; Ramlawi, B.; Mahmu d, F.; Kjelstrom, S.; Wertan, M.; Sutter, F. Robotic-Assisted Versu s Traditional Full-Sternotomy Coronary Artery Bypass Grafting Procedures: A Propensity-Matched Analysis of Hospital Costs. *A m. J. Cardiol.* **2024**, *213*, 12–19. https://doi.org/10.1016/j.amjcard. 2023.10.083.
29. Alemozaffar, M.; Chang, S. L.; Kacker, R.; Sun, M.; DeWolf, W. C.; Wagner, A. A. Comparing Costs of Robotic, Laparoscopic, an d Open Partial Nephrectomy. *J. Endourol.* **2013**, *27* (5), 560–565. https://doi.org/10.1089/end.2012.0462.
30. Paul, S.; Isaacs, A. J.; Jalbert, J.; Osakwe, N. C.; Salemi, A.; Girar di, L. N.; Sedrakyan, A. A Population-Based Analysis of Robotic -Assisted Mitral Valve Repair. *Ann. Thorac. Surg*. **2015**, *99* (5), 15 46–1553. https://doi.org/10.1016/j.athoracsur.2014.12.043.
31. Suri, R. M.; Thompson, J. E.; Burkhart, H. M.; Huebner, M.; Bo rah, B. J.; Li, Z.; Michelena, H. I.; Visscher, S. L.; Roger, V. L.; Da ly, R. C.; Cook, D. J.; Enriquez-Sarano, M.; Schaff, H. V. Improv ing Affordability Through Innovation in the Surgical Treatment of Mitral Valve Disease. *Mayo Clin. Proc.* **2013**, *88* (10), 1075-108 4. https://doi.org/10.1016/j.mayocp.2013.06.022.
32. Aljamaan, F.; Temsah, M.-H.; Altamimi, I.; Al-Eyadhy, A.; Jama l, A.; Alhasan, K.; Mesallam, T. A.; Farahat, M.; Malki, K. H. Refe rence Hallucination Score for Medical Artificial Intelligence Cha tbots: Development and Usability Study. *JMIR Med. Inform*. **202 4**, *12* (1), e54345. https://doi.org/10.2196/54345.
33. Brown, T. B.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.; Dhari wal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; Agarwa l, S.; Herbert-Voss, A.; Krueger, G.; Henighan, T.; Child, R.; Ram esh, A.; Ziegler, D. M.; Wu, J.; Winter, C.; Hesse, C.; Chen, M.; S igler, E.; Litwin, M.; Gray, S.; Chess, B.; Clark, J.; Berner, C.; Mc Candlish, S.; Radford, A.; Sutskever, I.; Amodei, D. Language M odels Are Few-Shot Learners. arXiv July 22, 2020. https://doi.org /10.48550/arXiv.2005.14165.
34. Kaddour, J.; Harris, J.; Mozes, M.; Bradley, H.; Raileanu, R.; Mc Hardy, R. Challenges and Applications of Large Language Model s. arXiv July 19, 2023. http://arxiv.org/abs/2307.10169 (accessed 2024-08-20).
35. Radford, A.; Narasimhan, K.; Salimans, T.; Sutskever, I. Improvi ng Language Understanding by Generative Pre-Training.
36. Zaman, M. ChatGPT for Healthcare Sector: SWOT Analysis. *Int. J. Res. Ind. Eng.* **2023**, *12* (3), 221–233. https://doi.org/10.221 05/riej.2023.391536.1373.
37. Clark, S. C. Can ChatGPT Transform Cardiac Surgery and Hea rt Transplantation? *J. Cardiothorac. Surg.* **2024**, *19* (1), 108. https:// doi.org/10.1186/s13019-024-02541-0.
38. Bektaş, M.; Pereira, J. K.; Daams, F.; van der Peet, D. L. ChatGPT in Surgery: A Revolutionary Innovation? *Surg. Today* **2024**, *54* (8), 964–971. https://doi.org/10.1007/s00595-024-02800-6.
39. Salihu, A.; Meier, D.; Noirclerc, N.; Skalidis, I.; Mauler-Wittwer, S.; Recordon, F.; Kirsch, M.; Roguelov, C.; Berger, A.; Sun, X.; Ab be, E.; Marcucci, C.; Rancati, V.; Rosner, L.; Scala, E.; Rotzinger, D. C.; Humbert, M.; Muller, O.; Lu, H.; Fournier, S. A Study of ChatGPT in Facilitating Heart Team Decisions on Severe Aortic Stenosis. *EuroIntervention* **2024**, *2*0 (8), e496–e503. https://doi.or g/10.4244/EIJ-D-23-00643.
40. Haug, C. J.; Drazen, J. M. Artificial Intelligence and Machine Le arning in Clinical Medicine, 2023. *N. Engl. J. Med.* **2023**, *388* (13), 1201–1208. https://doi.org/10.1056/NEJMra2302038.
41. Mann, D. L. Artificial Intelligence Discusses the Role of Artifici al Intelligence in Translational Medicine: A JACC: Basic to Trans lational Science Interview With ChatGPT. *JACC Basic Transl. Sci* . **2023**, *8* (2), 221–223. https://doi.org/10.1016/j. jacbts.2023.01.0 01.
42. Javaid, M.; Haleem, A.; Singh, R. P. ChatGPT for Health care S ervices: An Emerging Stage for an Innovative Perspective. Bench Council Trans. *Benchmarks Stand. Eval.* **2023**, *3* (1), 100105. http s://doi.org/10.1016/j.tbench.2023.100105.
43. *What is Fine-Tuning? | IBM*. https://www.ibm.com/topics/fine-tuning (accessed 2024-11-03).
44. Cao, S.; Fu, D.; Yang, X.; Wermter, S.; Liu, X.; Wu, H. Can AI Detect Pain and Express Pain Empathy? A Review from Emotio n Recognition and a Human-Centered AI Perspective. arXiv Dec ember 1, 2022. https://doi.org/10.48550/arXiv.2110.04249.
45. *CAIR launches CARES Copilot 1.0 LLM for surgery | Health Teach Asia*. https://healthtechasia.co/cair-launches-cares-copilat-1-0-1 lm-for-surgery/ (accessed 2024-08-24).
46. China Begins Testing AI Chatbot for Brain Surgeons in Hospita ls. *Bloomberg.com.* March 11, 2024. https://www.bloomberg.com/ news/articles/2024-03-11/china-begins-testing-ai-chatbot-for-b rain-surgeons-in-hospitals (accessed 2024-08-24).
47. *China Unveils Advanced AI Model for Medicine, Boosting Clinical Diagnosis----Chinese Academy of Sciences*. https://english.cas.cn/n ewsroom/cas_media/202403/t20240315_658496.shtml (accessed 2024-08-24).
48. Chang, Y.; Wang, X.; Wang, J.; Wu, Y.; Yang, L.; Zhu, K.; Chen, H.; Yi, X.; Wang, C.; Wang, Y.; Ye, W.; Zhang, Y.; Chang, Y.; Yu, P. S.; Yang, Q.; Xie, X. A Survey on Evaluation of Large Language Models. *ACM Trans Intell Syst Technol* **2024**, *15* (3), 39:1-39:45. h ttps://doi.org/10.1145/3641289.
49. *CARES Copilot*. https://www.cares-copilot.com/ (accessed 2024-08-24).
50. *AI Empowers Surgery: CAIR Unveils Large Multimodality Model f r Surgery----Chinese Academy of Sciences*. https://english.cas.cn/ne wsroom/news/202403/t20240315_658493.shtml (accessed 2024-08-24).
51. *Centre for Artificial Intelligence and Robotics (CAIR) Hong Kong I nstitute of Science & Innovation, Chinese Academy of Sciences*. https:/ /www.cair-cas.org.hk/articles/ai (accessed 2024-11-03).
52. Dave, T.; Athaluri, S. A.; Singh, S. ChatGPT in Medicine: An O verview of Its Applications, Advantages, Limitations, Future Pros pects, and Ethical Considerations. *Front. Artif. Intell.* **2023**, *6*.http s://doi.org/10.3389/frai.2023.1169595.
53. Salomon, I.; Olivier, S. Artificial Intelligence in Medicine: Adva ntages and Disadvantages for Today and the Future. *Int. J. Surg. Open* **2024**, *62* (4), 471. https://doi.org/10.1097/IO9.0000000000 00133.
54. Maleki, N.; Padmanabhan, B.; Dutta, K. AI Hallucinations A M isnomer Worth Clarifying. In *2024 IEEE Conference on Artificial Intelligence (CAI)*; IEEE: Singapore, Singapore, 2024; pp 133–13 8. https://doi.org/10.1109/CAI59869.2024.00033.
55. Milmo, D. Two US Lawyers Fined for Submitting Fake Court C

itations from ChatGPT. *The Guardian*. June 23, 2023. https://www.theguardian.com/technology/2023/jun/23/two-us-lawyer-fined-submitting-fake-court-citations-chatgpt (accessed 2024-11-14)
56. *Lawyers submitted bogus case law created by ChatGPT. A judge fined them $5,000*. AP News. https://apnews.com/article/artificial-intelligence-chatgpt-fake-case-lawyers-d6ae9fa79d0542db9e1455397aef381c (accessed 2024-11-14).
57. *HHEM 2.1*. Vectara. https://www.vectara.com/blog/hhem-2-1-a-better-hallucination-detection-model (accessed 2024-11-11).
58. Wang, P.; Wang, Z.; Li, Z.; Gao, Y.; Yin, B.; Ren, X. SCOTT: Self-Consistent Chain-of-Thought Distillation. arXiv August 30, 2023. https://doi.org/10.48550/arXiv.2305.01879.
59. Suzuoki, S.; Hatano, K. Reducing Hallucinations in Large Language Models: A Consensus Voting Approach Using Mixture of Experts.
60. Li, J.; Chen, J.; Ren, R.; Cheng, X.; Zhao, W. X.; Nie, J.-Y.; Wen, J.-R. The Dawn After the Dark: An Empirical Study on Factuality Hallucination in Large Language Models. arXiv January 6, 2024. http://arxiv.org/abs/2401.03205 (accessed 2024-11-06).
61. Gekhman, Z.; Yona, G.; Aharoni, R.; Eyal, M.; Feder, A.; Reichart, R.; Herzig, J. Does Fine-Tuning LLMs on New Knowledge Encourage Hallucinations? arXiv October 1, 2024. https://doi.org/10.48550/arXiv.2405.05904.
62. Rumiantsau, M.; Vertsel, A.; Hrytsuk, I.; Ballah, I. Beyond Fine-Tuning: Effective Strategies for Mitigating Hallucinations in Large Language Models for Data Analytics. arXiv October 26, 2024. https://doi.org/10.48550/arXiv.2410.20024.
63. Xu, Z.; Jain, S.; Kankanhalli, M. Hallucination Is Inevitable: An Innate Limitation of Large Language Models. arXiv January 22, 2024. https://doi.org/10.48550/arXiv.2401.11817.
64. Arcaya, M. C.; Arcaya, A. L.; Subramanian, S. V. Inequalities in Health: Definitions, Concepts, and Theories. *Glob. Health Action* **2015**, *8*, 27106. https://doi.org/10.3402/gha.v8.27106.
65. Tat, E.; Bhatt, D. L.; Rabbat, M. G. Addressing Bias: Artificial Intelligence in Cardiovascular Medicine. Lancet Digit. *Health* **2020**, *2* (12), e635–e636. https://doi.org/10.1016/S2589-7500(20)30249-1.
66. Health, C. for D. ad R. Good Machine Learning Practice for Medical Device Development: Guiding Principles. *FDA* **2023**.
67. Qraitem, M.; Saenko, K.; Plummer, B. A. Bias Mimicking: A Simple Sampling Approach for Bias Mitigation. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: Vancouver, BC, Canada, 2023; pp 20311–20320. https://doi.org/10.1109/CVPR52729.2023.01945.
68. Buda, M.; Maki, A.; Mazurowski, M. A. A Systematic Study of the Class Imbalance Problem in Convolutional Neural Networks. *Neural Netw.* **2018**, *106*, 249–259. https://doi.org/10.1016/j.neunet.2018.07.011.
69. Health, C. for D. and R. *Software as a Medical Device (SAMD): Clinical Evaluation*. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/software-medical-device-samd-clinical-evaluation (accessed 2024-11-11).
70. Yang, J.; Soltan, A. A. S.; Eyre, D. W.; Yang, Y.; Clifton, D. A. An Adversarial Training Framework for Mitigating Algorithmic Biases in Clinical Machine Learning. *Npj Digit. Med.* **2023**, *6* (1), 1-10. https://doi.org/10.1038/s41746-023-00805-y.
71. Yang, Y.; Lin, M.; Zhao, H.; Peng, Y.; Huang, F.; Lu, Z. A Survey of Recent Methods for Addressing AI Fairness and Bias in Biomedicine. *J. Biomed. Inform.* **2024**, *154*, 104646. https://doi.org/10.1016/j.jbi.2024.104646.
72. Abràmoff, M. D.; Tarver, M. E.; Loyo-Berrios, N.; Trujillo, S.; Char, D.; Obermeyer, Z.; Eydelman, M. B.; Foundational Principles of Ophthalmic Imaging and Algorithmic Interpretation Working Group of the Collaborative Community for Ophthalmic Imaging Foundation, W.; Maisel, W. H. Considerations for Adressing Bias in Artificial Intelligence for Health Equity. *NPJ Digit. Med.* **2023**, *6*, 170. https://doi.org/10.1038/s41746-023-00913-9.
73. Jan 26, C. G. / P.; 2022. *Robot performs first laparoscopic surgery without human help*. The Hub. https://hub.jhu.edu/2022/01/26/star-robot-performs-intestinal-surgery/ (accessed 2024-08-24).

■ Author

Yonas Getenet Bekele is a high school student at Menelik II Secondary School in Addis Ababa, Ethiopia, with an extraordinary passion for computer science, AI, engineering, and biology. For him, research and exploration are the ways he can learn new things and use them as tools to solve interdisciplinary problems.

DOI: 10.36838/v7i3.97