



Writing the paper “*Unveiling artificial intelligence: an insight into ethics and applications in anesthesia*” implementing the large language model ChatGPT: a qualitative study

Marco Cascella^{1^}, Jonathan Montomoli², Valentina Bellini³, Alessandro Ottaiano⁴, Mariachiara Santorsola⁴, Francesco Perri⁵, Francesco Sabbatino⁶, Alessandro Vittori⁷, Elena Giovanna Bignami³

¹Department of Anesthesia and Critical Care, Istituto Nazionale Tumori IRCCS Fondazione G. Pascale, Naples, Italy; ²Department of Anesthesia and Intensive Care, Infermi Hospital, AUSL Romagna, Rimini, Italy; ³Division of Anesthesiology, Critical Care and Pain Medicine, Department of Medicine and Surgery, University of Parma, Parma, Italy; ⁴SSD Innovative Therapies for Abdominal Metastases, Abdominal Oncology, Istituto Nazionale Tumori IRCCS Fondazione G. Pascale, Naples, Italy; ⁵Medical and Experimental Head and Neck Oncology Unit, Istituto Nazionale Tumori IRCCS Fondazione G. Pascale, Naples, Italy; ⁶Oncology Unit, Department of Medicine, Surgery and Dentistry, University of Salerno, Salerno, Italy; ⁷Department of Anesthesia and Critical Care, ARCO Roma, Ospedale Pediatrico Bambino Gesù, IRCCS, Rome, Italy

Contributions: (I) Conception and design: M Cascella; (II) Administrative support: F Sabbatino; (III) Provision of study materials or patients: M Cascella; (IV) Collection and assembly of data: M Cascella, V Bellini, J Montomoli, A Vittori, EG Bignami; (V) Data analysis and interpretation: M Santorsola, A Ottaiano, F Sabbatino, A Vittori; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Marco Cascella, MD. Department of Anesthesia and Critical Care, Istituto Nazionale Tumori IRCCS Fondazione G. Pascale, Via Mariano Semmola, 53, 80100 Naples, Italy. Email: m.cascella@istitutotumori.na.it.

Background: This text “*Unveiling artificial intelligence: an insight into ethics and applications in anesthesia*” addresses the potential applications and limits of artificial intelligence (AI) in anesthesia. The written content was produced with the help of ChatGPT. It is a large language model (LLM) capable of performing a variety of natural language processing (NLP) tasks. The aim of this paper is to evaluate the ability and limits of the chatbot as an aid in writing a scientific text.

Methods: In this qualitative study, five domain experts (MC, JM, VB, AV, and EGB) elaborated on the questions for the chatbot and integrated the answers based on the knowledge of the available scientific literature. The output was incorporated, edited, and revised based on the knowledge of the available scientific literature. The first author (MC) assessed the effectiveness of the output and its applicability in writing the paper, as a percentage of words used. All the authors expressed an opinion on the quality of the text and highlighted the pros and cons of using the technology.

Results: In the final draft of the text, the scientific output from ChatGPT has been significantly revised and integrated. Many important inaccuracies were found.

Conclusions: The utilization of LLMs like ChatGPT presents notable limitations. However, by effectively harnessing specific attributes, these tools can provide advantages for specific tasks. Consequently, it becomes crucial to establish guidelines for their application in scientific writing.

Keywords: Artificial intelligence (AI); anesthesia; ethics; chatbot; qualitative study

Received: 03 February 2023; Accepted: 15 June 2023; Published online: 30 June 2023.

doi: 10.21037/jmai-23-13

View this article at: <https://dx.doi.org/10.21037/jmai-23-13>

[^] ORCID: 0000-0002-5236-3132.

Introduction

Large language models (LLMs) are advanced artificial intelligence (AI) systems designed to mimic human language processing abilities. They use deep learning techniques, such as neural networks, to analyze vast amounts of text data from various sources, including books, articles, and websites, to generate highly coherent and realistic text. LLMs find and structure patterns and relationships within the training data to predict the most likely words or phrases in a specific context (1). This ability to comprehend and generate language is beneficial in various fields of natural language processing (NLP), such as machine translation and text generation.

A widespread type of LLM is the generative pre-trained transformer (GPT), developed by OpenAI (San Francisco, CA, USA) in 2018. The first iteration of GPT was trained on a 40-GB text dataset using a variant of the transformer architecture and had a model size of 1.5 billion parameters. In 2020, OpenAI released GPT-3, a more advanced version of GPT, which was trained on a massive 570-GB text dataset and had a model size of 175 billion parameters, making it one of the most powerful language models currently. Notably, due to its ability to be fine-tuned for specific tasks, GPT-3 can write essays, compose poetry, or codices using a smaller amount of task-specific data. The newest version, GPT-4, was released in March 2023. It can enhance the potential of technology to support various medical tasks. These advancements might include improved reliability and the ability to analyze and interpret text within images.

ChatGPT, a variant of GPT-3, was specifically developed for conversational AI and can engage in dialogue with users. The underlying transformer model, a deep neural network architecture, uses self-attention to weigh the importance of different parts of the input data. Pretrained on a large corpus of text, it learns context and meaning from sequential data. Therefore, ChatGPT generates natural language responses based on user inputs, utilizing its understanding, and learned patterns. It breaks down input into tokens, processes them with the transformer model, and generates output iteratively. In particular, a token refers to a unit of text (e.g., a word or a character) that is treated as a single entity by the NLP model. In the context of ChatGPT, the iterative process persists, as the model generates new tokens and output text at each step until a predefined stopping criterion is satisfied.

Consequently, a revolution is underway in the field of writing (2). However, it is advisable to know the applications and the limits of the technology, referring above all to the ethical aspects (3).

On these premises, we present a text entitled: “*Unveiling artificial intelligence: an insight into ethics and applications in anesthesia*”. This manuscript is an “experiment” in writing a text “in collaboration” with the machine.

Research question

This study poses the research question: what is the accuracy and reliability of a text on AI and anesthesia written with the assistance of ChatGPT?

Methods

This study is qualitative, with the purpose of describing the level of reliability of a text written using a hybrid approach between domain experts and LLM.

Five domain experts (MC, VM, JM, AV, and EGB) elaborated on the topics and formulated the questions submitted to the chatbot. The first author (MC) assessed the effectiveness of the output through a checklist and, finally, evaluated its applicability in writing the final paper, expressed as a percentage of words used. The output was integrated, edited (including revision of grammatical/punctuation errors), and revised based on the knowledge of the available scientific literature.

Subsequently, all the authors provided a comprehensive evaluation of the text and discussed its strengths and weaknesses in regard to the application of the technology.

Highlight box

Key findings

- The scientific output from ChatGPT, although generated through advanced AI algorithms, can contain significant inaccuracies and errors.

What is known and what is new?

- Limited knowledge cutoff, lack of context, or oversimplification of complex scientific concepts are the main limitations.
- ChatGPT can be useful for notional purposes (e.g., definitions, and historical contexts) as well as for simple language editing (e.g., synonyms search).

What is the implication, and what should change now?

- It is crucial to set guidelines for the utilization of LLM tools such as ChatGPT in scientific writing, as they come with significant limitations.

Statistical analysis

Given the qualitative nature of the analysis, statistical tools were not employed in the study. Instead, a thematic analysis approach was utilized to thoroughly examine and interpret the data. This method involved a systematic exploration of the qualitative data (textual documents), to identify recurring themes, patterns, and underlying meanings.

Results

The text generated by ChatGPT after editing

Despite editing, it may contain several inaccuracies. For this reason, for example, the references produced by the chatbot are not reported. More importantly, the text should not be viewed as a credible and practical scientific representation of the subject, but rather as a linguistic experiment and a test of feasibility.

The following is the revised version of the manuscript titled “*Unveiling artificial intelligence: an insight into ethics and applications in anesthesia*”. It has been edited and structured into several sections, including an “Introduction”, narrative paragraphs addressing “Ethical and social issues in AI” and “AI in anesthesia”, and a concluding section.

Introduction

The term “artificial intelligence (AI)” was first used by John McCarthy, a computer scientist, in 1956 during a conference held in the summer of 1956 at Dartmouth College in Hanover, New Hampshire. It was a significant event in the history of AI. During the meeting, McCarthy, Marvin Minsky, Nathaniel Rochester, Claude Shannon, and the attendees used the term “artificial intelligence (AI)” to describe a field of study focused on creating machines that can exhibit intelligent behavior and perform tasks that would normally require human intelligence. The Dartmouth Conference is considered the birth of AI as a field of study, and it laid the foundations for the development of AI technology in the years that followed.

Another founder of the discipline is Alan Turing. He was a mathematician and computer scientist who is widely considered to be the father of theoretical computer science. Six years before the conference, in 1950, Turing published an article in the journal *Mind* titled “*Computing machinery and intelligence*”, in which he introduced the concept of the “Turing Test” as a way to determine whether a machine can exhibit human-like intelligent behavior that is indistinguishable from a human. In the article, Turing

argued that if a machine can carry on a conversation with a human in such a way that the human is unable to distinguish it from another human, then it can be considered intelligent. Turing’s ideas about AI and the Turing Test have had a significant impact on the field of computer science and have influenced the development of the AI discipline.

Beyond the historical notes, AI refers to the ability of a computer or machine to perform tasks that would normally require human intelligence, such as learning, problem-solving, and decision-making. Ultimately, the objective of AI is to develop systems capable of executing tasks that typically require human intelligence, with the aim of enhancing efficiency and precision.

AI subtypes

There are several different types of AI, including rule-based AI, machine learning (ML), neural network, and expert system AI.

Rule-based AI

This type of AI follows a set of predetermined rules to perform a task. It is based on the idea of creating a set of explicit rules that a machine can follow to solve a problem. These subtypes of AI are relatively simple and are often used to perform specific tasks where the rules are well-defined, and the problem domain is limited.

In medicine, a rule-based AI system designed to diagnose medical conditions might be programmed with a set of rules that describe the symptoms of different diseases. For example, the system might be programmed with the rule: “If a patient has a fever and a cough, then they may have the flu”. The system would then use this rule, along with other rules about the symptoms of different diseases, to diagnose a patient’s condition based on their symptoms.

ML

This type of AI uses algorithms to learn from data and improve its performance over time. It is based on the idea of creating algorithms that can automatically learn and improve from experience, without being explicitly programmed.

There are several different types of ML algorithms that can be classified into two main categories: supervised and unsupervised. In supervised learning, the algorithm is trained on a dataset that has been labeled with the correct output or result. This allows the algorithm to learn how to map inputs to outputs, and it can then make predictions or decisions on new data based on its learning. Examples of supervised learning algorithms include linear regression, logistic regression, and support vector machines.

Unsupervised learning involves the use of an algorithm that is not given any labeled data. Instead, the algorithm must find patterns and structure in the data on its own. This type of ML is often used for tasks such as clustering, in which the goal is to group similar items together. The most used unsupervised learning algorithms include k-means clustering and principal component analysis.

There are also other subtypes of ML algorithms, such as semi-supervised learning, reinforcement learning, and deep learning. These algorithms are typically used for more complex tasks and involve advanced techniques and approaches.

In medicine, ML is used for diagnosis, predictive modeling, drug discovery, and clinical decision support. For diagnostic aims, ML algorithms can be trained on medical data to recognize patterns and make accurate diagnoses. For example, a ML algorithm trained on medical images could be used to identify cancerous tumors. ML algorithms can be also used to develop predictive models. They can analyze patient data and make predictions about the likelihood of certain outcomes, such as the risk of developing a particular disease or the likelihood of a patient responding to a particular treatment. Drug discovery is an important application of ML. These strategies can be implemented to analyze chemical compounds and predict their potential as drugs, helping to speed up the drug discovery process. Finally, ML can provide recommendations and assist in decision-making for clinicians, helping to improve patient care and reduce errors.

Neural networks

This type of AI is inspired by the way the human brain works and is composed of interconnected “neurons” that process and transmit information. It is based on the idea of creating a network of artificial neurons that can process and transmit information in a way that is like the human brain.

Each neural network is composed of a variable number of interconnected processing nodes, or “neurons”, which are organized into layers. The input layer receives data, which is then processed and transmitted through one or more hidden layers before reaching the output layer. Each layer of neurons processes the data and passes it on to the next layer, with the goal of recognizing patterns and relationships in the data. The connections between neurons are weighted, which means that each connection has a specific value that determines the strength of the connection. During the training process, the weights of the connections are adjusted to improve the accuracy of the neural network. Several types of neural networks have been

developed. They can be broadly classified based on their architecture and the type of problem they are designed to solve (*Figure 1*). Some common types of neural networks include feedforward neural networks (the data flows in one direction from the input layer to the output layer without looping back) used for classification and prediction, convolutional neural networks (CNNs) to process data with a grid-like topology, such as images, recurrent neural networks (RNNs) to process sequential data, such as time series or natural language (e.g., for language translation), and autoencoders used for dimensionality reduction and feature learning (encoder and decoder elements work together to learn a compressed representation of the data). There are also many other types of neural networks and variations on these basic architectures, such as long short-term memory (LSTM) networks and generative adversarial networks (GANs).

One of the key advantages of neural networks is their ability to learn and adapt to new data. As the neural network is exposed to more data, it can improve its performance and accuracy by adjusting the weights of the connections between neurons.

Neural networks are often used for tasks such as image and speech recognition, language translation, and forecasting. They are particularly well-suited for tasks that involve complex patterns and relationships in the data. However, they can also be computationally intensive and require large amounts of data for training.

Expert system AI

It is designed to mimic the decision-making abilities of a human expert in a particular field. Expert AI is based on the idea of creating a system that can use knowledge and reasoning to solve problems and make decisions in a way that is similar to a human expert. Thus, expert system AI is often implemented using a “knowledge base” that contains information about a particular domain, such as medical knowledge or technical knowledge. The system is then able to use this information to reason about problems and make decisions. These approaches have been used in a variety of fields, including medicine, finance, and customer service.

Some common algorithms used in expert system AI include rule-based algorithms (they use a set of explicit rules), case-based reasoning algorithms (past examples or “cases” to solve new problems), decision tree algorithms (tree-like structure to represent the possible decisions and outcomes of a problem), and Bayesian networks (represent the uncertainties and dependencies of a problem).

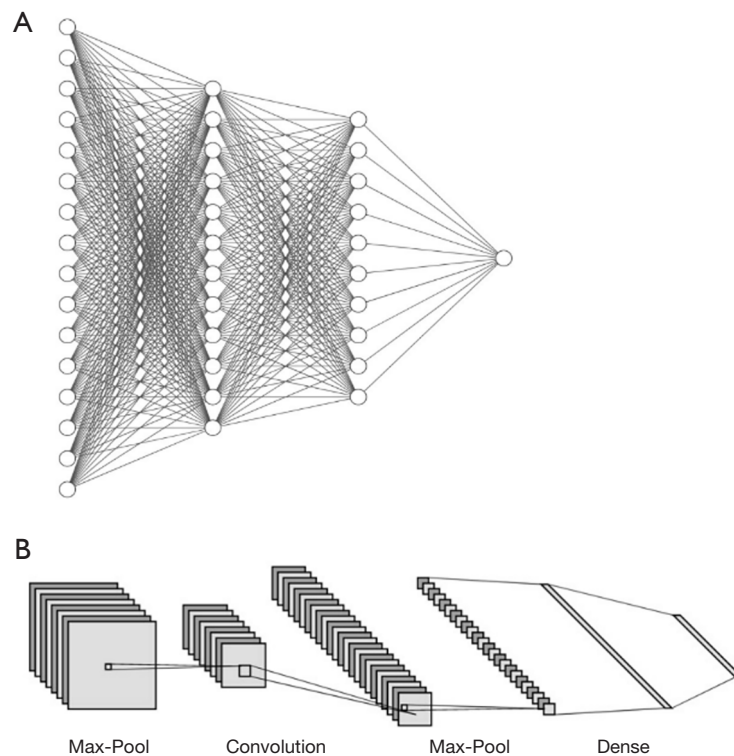


Figure 1 Examples of neural network architectures. (A) A feedforward neural network with an input layer of 16 nodes, two hidden layers, and a single-node output layer. (B) LeNet is a simple CNN architecture that consists of a series of convolutional and pooling layers, followed by a few fully connected (dense) layers. Figures developed from NNSVG (<https://alexlenail.me/NN-SVG/LeNet.html>). Prompt: “Indicate a website for designing LeNet convolutional neural network architecture”. CNN, convolutional neural network.

Ethical and social issues in AI

Since AI is a rapidly developing field, many ethical and social issues need to be considered as it continues to advance, especially in fields such as medicine. For example, AI systems may be biased against certain groups of people, either intentionally or unintentionally. This can occur if the data used to train the AI system or the algorithms adopted to develop the AI system are unbalanced or inaccurate. Moreover, there are concerns about the potential for AI systems to infringe upon privacy, either by collecting and storing personal data or by using that data in ways that are not transparent or controlled by the individual. Another issue is AI autonomy. Notably, there are concerns about the increasing reliance on AI strategies and the potential for these systems to make decisions without human oversight. This raises questions about the accountability and responsibility of AI systems and their potential to infringe upon human autonomy. Other problems concern unemployment and transparency. Probably, the increasing use of AI may lead to job displacement and unemployment,

as machines and algorithms may be able to perform certain tasks more efficiently than humans. Finally, a special issue is the “black box” nature of many AI systems as it can make it difficult to understand how they make decisions and to hold them accountable for their actions. Explainable AI (XAI) is a field of AI that focuses on developing systems that can provide explanations for their actions, predictions, or decisions. It can help to build trust in AI systems, especially when they are used in sensitive applications such as healthcare, finance, or criminal justice. Consequently, there is a need for appropriate regulation of AI to ensure that it is developed and used ethically and responsibly.

AI in anesthesia

Remarkably, AI has the potential to revolutionize the field of anesthesiology by improving patient care and safety. There are several ways in which AI can be used in anesthesiology. For example, for guiding anesthetic drug dosing, perioperative monitoring, and predictive modeling, as well as education and training.

Anesthetic drug dosing

Anesthetic drug dosing is a complex process that requires careful consideration of a number of factors, including a patient's age, weight, medical history, and the type of surgery being performed. Misjudging the dosage can have serious consequences, including inadequate anesthesia, which can result in patient discomfort and increased risk of complications, or oversedation, which can lead to potentially fatal outcomes.

AI has the potential to assist in the precise dosing of anesthetic drugs by using algorithms to analyze patient data and provide recommendations for dosage. A number of studies have demonstrated the effectiveness of AI in this regard. For example, it was demonstrated that an AI-based system was able to accurately predict the optimal propofol dosage for individual patients during surgery, leading to improved patient outcomes and reduced risk of adverse events. In another investigation, the authors developed and validated an ML model based on a decision tree algorithm for predicting the requirement of propofol during surgery. The model was trained on data from a large number of patients and was able to accurately predict the optimal propofol dosage for individual patients. Intraoperative opioid administration was also investigated, and the authors developed and validated a model based on a random forest algorithm for predicting the intraoperative requirements of propofol and remifentanyl.

Monitoring

Another potential application is monitoring. AI-powered monitoring systems can continuously analyze patient data and alert anesthesiologists to potential problems. These systems can also provide recommendations for treatment and can help to reduce the risk of complications.

There have been a number of studies that have explored the use of AI for anesthesia monitoring. For example, some researchers have developed ML algorithms that can analyze data from various anesthesia monitors (e.g., electrocardiogram, blood pressure, etc.) to identify patterns that may indicate a change in a patient's condition. Other studies have focused on using NLP to automatically extract information from electronic medical records and other sources to aid in decision-making during anesthesia.

Predictive models for outcomes

AI is also employed for the development of predictive models: AI algorithms can analyze patient data and predict the likelihood of certain outcomes, such as the need for additional anesthetic drugs or the risk of postoperative complications. This can help anesthesiologists to make

more informed decisions about patient care.

One study used ML algorithms to develop predictive models for postoperative nausea and vomiting (PONV) based on patient characteristics and anesthetic factors. The model was able to accurately predict the likelihood of PONV in patients, allowing anesthesiologists to tailor their treatment accordingly. Other investigations focused on predictive modeling for postoperative pain management. For example, in a retrospective analysis of 2,200 patients, it was found that an ML algorithm was able to predict the likelihood of patients experiencing moderate to severe pain after surgery based on patient characteristics and surgical factors. According to the authors, the predictive model developed in the study could potentially be used to proactively manage pain and improve patient outcomes.

In a systematic review and meta-analysis, the authors analyzed the available evidence on the use of AI in anesthesia. The review found that AI has the potential to improve patient safety and outcomes, but also highlighted the need for further research to assess the effectiveness and safety of AI in this field. In a more recent evidence-based analysis, the authors investigated different postoperative outcomes, including mortality risk, systemic complications (e.g., pulmonary, cardiovascular, acute kidney injury), intensive care unit admission, anesthesiologic risk, and prolonged length of hospital stay. They reported that the most frequently used algorithms were gradient boosting, random forest, logistic regression, artificial neural networks, and support vector machines. The models with the best performance were random forest and gradient boosting [area under the curve (AUC) >0.90]. The authors found that using ML algorithms can result in more accurate prediction of outcomes compared to validated prognostic scores and traditional statistical methods.

Education and training

Applications of AI can be used to create virtual simulations and training programs that can help anesthesiologists to improve their skills and knowledge. Different studies demonstrated the potential of AI to assist in the assessment of trainee performance in anesthesia. For instance, it was developed and validated a ML model for automated assessment of trainee anesthetic performance. The model was trained on data from a large number of trainees and was able to accurately predict the performance of individual participants based on their recorded data. Furthermore, in their systematic review, the authors analyzed the available evidence on the use of AI in anesthesia education and proved that AI has the potential to improve the efficiency

Table 1 Prompts and output evaluation

Prompt	Notes on outputs	Writer interventions
Limitations of ChatGPT	Relevant and clearly expressed output	Logical sequence organization
The history of AI	Relevant and clearly expressed output	Logical sequence organization
AI subtypes	Too succinct and not very detailed	Re-writing
	Not useful information	
Example of rule-based AI	Pertinent output	Sentences review and editing
Ethical and social issues in AI	Generic response	Re-writing
	Not useful information	
Give references for the previous query [†]	Incorrect references	Re-writing
	Not useful information	
AI in anesthesia	Not complete list	Re-writing
	Partially useful information	
Write a conclusion from the text [‡]	A brief summary	Editing

[†], examples of references in [Appendix 1](#); [‡], produced from the prompts “Ethical and social issues in AI” and “AI in anesthesia” and revision. AI, artificial intelligence.

and effectiveness of training and assessment in anesthesia, but also highlighted the need for further research to assess the effectiveness and safety of AI in this context.

Conclusion

The use of AI in anesthesiology has the potential to significantly improve patient care and safety. However, it is important for anesthesiologists to carefully consider the ethical implications of using AI in medical practice and to ensure that these systems are used responsibly.

Evaluation of the text generated by ChatGPT

The final text is 2,648 words. The first author (MC) organized the sessions and used the output he deemed necessary. The individual sentences of the output have been checked for scientific relevance. The answers to the prompts (output) were evaluated (*Table 1*). Prompts and outputs are reported in [Appendix 2](#), checklist for text evaluation is in [Appendix 3](#).

A variable percentage of the text produced by ChatGPT was used for the final draft of the text (*Table 2*). The criteria used included relevance, completeness, consistency, and logical coherence (logical progression and cohesive structure), verification of information accuracy (reliable information), absence of redundancy (unnecessary repetition of information or ideas within the text), and clarity and readability.

Discussion

In this paper, we have evaluated the possible use of a chatbot for scientific writing. Overall, ChatGPT's help has been marginal. As a result, there are several factors that need to be considered.

It is essential for the scientific community to fully comprehend the limitations and capabilities of LLMs. This involves identifying specific tasks and areas where they can be useful, as well as any potential challenges and limitations they may have (4).

It is important to emphasize that chatbots cannot replace humans in the realm of scientific writing (5). As Thorp (6) affirmed, “ChatGPT is fun, but not an author”. For this purpose, it is essential to be able to define the limits and strategies to identify plagiarism. The answers and outputs are often difficult to identify as constructed. Recently, Gao *et al.* (7) published an interesting study in which the chatbot generated 50 medical research abstracts based on articles published in medical journals such as *JAMA*, *New England Journal of Medicine*, *BMJ*, *Lancet*, and *Nature Medicine*. The researchers then evaluated the generated abstracts by using a plagiarism checker and AI detection tool. Surprisingly, hardly any plagiarism was identified. For this aim, several tools for detecting AI-based plagiarism are being evaluated (8).

Concerning reliability, several limitations of ChatGPT

Table 2 ChatGPT output used for the final text

Paper section	Percentage of ChatGPT text used	Notes and editing
Introduction	<30%	GPT-3 and ChatGPT features Extensive editing
Methods	0	–
Results (article text)		
Introduction	30%	AI subtypes description
Ethical and social issues in AI	<10%	Extensive editing
AI in anesthesia	<10%	Extensive editing
Conclusion	>50%	Editing of a text from the chatbot
Tables/figures	0	–
Conclusions	>70%	Text editing
References	0	Not accurate data provided [†]

[†], the DOI precisely corresponds solely to the journal title (see [Appendix 1](#)). GPT, generative pre-trained transformer; AI, artificial intelligence; DOI, digital object identifier.

Table 3 ChatGPT limitations

Limitation	Rationale	Ref.
Bias	It is trained on a large dataset of internet text, which may contain biases and stereotypes. This can lead to biased or stereotypical output when generating text	(2,9)
Lack of common sense	It has only simple logical reasoning abilities. Although it has a vast amount of knowledge, the chatbot lacks the ability to understand and reason about the world in the same way that humans do. This can lead to nonsensical or misleading output when answering questions or generating text	(10-12)
Lack of creativity	It is a highly sophisticated model, but it is still a machine, and it can't generate new ideas in the way that a human can	(2,10)
Safety and ethical concerns	It is a powerful tool that can be used for malicious purposes, such as creating fake news or impersonating individuals online	(10)
Unintended consequences	Its ability to generate highly coherent and fluent text may lead to the proliferation of misinformation and manipulation of public opinion	(5,10)
Lack of update	ChatGPT knowledge is current up until 2021, and the model may have been updated or changed since then	(2,11)

must be highlighted (9-13) (*Table 3*).

The advantages and disadvantages listed below showcase the benefits and drawbacks of utilizing ChatGPT to assist with scientific writing.

Pro:

- ❖ The use of the ChatGPT has been particularly effective for notional purposes (e.g., definitions, and historical contexts);
- ❖ The tool was useful for simple language editing (e.g., synonyms search);

- ❖ On a broader scale, LLMs like ChatGPT can be beneficial in reviewing and synthesizing large amounts of data;
- ❖ These models can also be employed to manage and analyze complex data, as well as extract meaningful information from texts;
- ❖ ChatGPT can be useful to generate and organize subtopics (14).

Cons:

- ❖ The phenomenon known as “hallucination”

emerged. This is a widespread problem among many NLP models (15). In other words, ChatGPT generates answers that sound credible but may be incorrect or nonsensical. It needs to be understood and addressed;

- ❖ The output not aligning with the author's intended meaning or tone. It leads to the risk of producing irrelevant or inaccurate information;
- ❖ Important limitations concern its full understanding of the context and nuances of a specific scientific subject (16);
- ❖ Another significant challenge is that ChatGPT can perpetuate biases that exist within the training data. Thus, it is important to be mindful of these limitations and to carefully evaluate the outputs of these models before using them in real-world applications (17);
- ❖ Specific limitations are the lack of updating and the impossibility of accessing datasets and libraries. Nevertheless, this limit could be exceeded as a recent model was trained on the PubMed Abstract, at Stanford University. Further improvements are needed for enhancing its accuracy (18);
- ❖ The provided references are unreliable (Appendix 3).

Limitations of the study

This study has significant limitations that should be taken into consideration. The use of a text percentage value is limited. Qualitative assessment variables would provide a better explanation. A subsequent investigation, involving NLP experts, will focus on text analysis. Furthermore, a thorough examination of the chatbot's output regarding data construction is necessary. Additionally, it would have been more beneficial if the chatbot's responses were generated through a standardized set of sequential prompts and were tested for reproducibility through a semiquantitative strategy.

Conclusions

In this brief experiment, we have emphasized various advantages and disadvantages of utilizing a chatbot as a potential tool in medical writing. This exploration sheds light on the benefits, such as increased efficiency and accuracy, as well as the drawbacks, such as potential biases and limitations. By highlighting these pros and cons, we aim to provide a comprehensive overview of the potential advantages of using a chatbot in medical writing

and to encourage further examination and evaluation of this technology in this field. Nevertheless, pivotal ethical aspects such as plagiarism and data construction need to be addressed. Concerning the research question, as ChatGPT stated: "It is important to use this technology responsibly and to be aware of these limitations in order to make the most of its capabilities while minimizing potential risks".

Acknowledgments

We thank Valeria Vicario, Clinical Study Coordinator, and IT Engineer at the Istituto Nazionale Tumori Fondazione G. Pascale for scientific and bibliographic assistance.

Funding: None.

Footnote

Data Sharing Statement: Available at <https://jmai.amegroups.com/article/view/10.21037/jmai-23-13/dss>

Peer Review File: Available at <https://jmai.amegroups.com/article/view/10.21037/jmai-23-13/prf>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://jmai.amegroups.com/article/view/10.21037/jmai-23-13/coif>). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521:436-44.
2. Gordijn B, Have HT. ChatGPT: evolution or revolution?

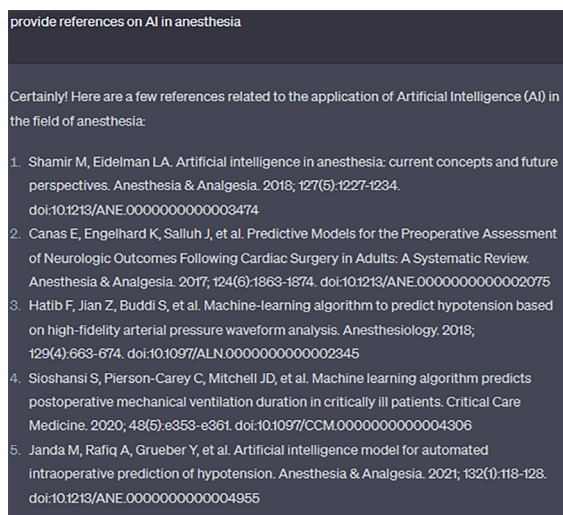
- Med Health Care Philos 2023;26:1-2.
3. Else H. Abstracts written by ChatGPT fool scientists. Nature 2023;613:423.
 4. Stokel-Walker C. ChatGPT listed as author on research papers: many scientists disapprove. Nature 2023;613:620-1.
 5. Looi MK. Sixty seconds on ... ChatGPT. BMJ 2023;380:205.
 6. Thorp HH. ChatGPT is fun, but not an author. Science 2023;379:313.
 7. Gao CA, Howard FM, Markov NS, et al. Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. NPJ Digit Med 2023;6:75.
 8. GPTZero. Available online: <https://gptzero.substack.com/> (Last Accessed: January 31, 2023).
 9. Sallam M. ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. Healthcare (Basel) 2023;11:887.
 10. Cascella M, Montomoli J, Bellini V, et al. Evaluating the Feasibility of ChatGPT in Healthcare: An Analysis of Multiple Clinical and Research Scenarios. J Med Syst 2023;47:33.
 11. Lorentzen T. Nuance is revolutionising the contact centre with GPT technology. 2023. Available online: <https://whatsnext.nuance.com/en-gb/customer-engagement/GPT-powered-capability-nuance-contact-centre/?print=pdf>
 12. Lee H. The rise of ChatGPT: Exploring its potential in medical education. Anat Sci Educ 2023. [Epub ahead of print]. doi: 10.1002/ase.2270.
 13. Sun GH, Hoelscher SH. The ChatGPT Storm and What Faculty Can Do. Nurse Educ 2023;48:119-24.
 14. Huang J, Tan M. The role of ChatGPT in scientific communication: writing better scientific review articles. Am J Cancer Res 2023;13:1148-54.
 15. Rohrbach A, Hendricks LA, Burns K, et al. Object hallucination in image captioning. arXiv preprint 2018. doi: 10.48550/arXiv.1809.02156.
 16. Temsah O, Khan SA, Chaiah Y, et al. Overview of Early ChatGPT's Presence in Medical Literature: Insights From a Hybrid Literature Review by ChatGPT and Human Experts. Cureus 2023;15:e37281.
 17. Graham F. Daily briefing: Will ChatGPT kill the essay assignment? Nature 2022. [Epub ahead of print]. doi: 10.1038/d41586-022-04437-2.
 18. PubMedGPT 2.7B. Available online: <https://crfm.stanford.edu/2022/12/15/pubmedgpt.html> (Last Accessed: January 31, 2023).

doi: 10.21037/jmai-23-13

Cite this article as: Cascella M, Montomoli J, Bellini V, Ottaiano A, Santorsola M, Perri F, Sabbatino F, Vittori A, Bignami EG. Writing the paper “*Unveiling artificial intelligence: an insight into ethics and applications in anesthesia*” implementing the large language model ChatGPT: a qualitative study. J Med Artif Intell 2023;6:9.

Appendix 1 (references from ChatGPT)

The precise DOIs match the following:



- (I) Bai XW. Anesthesia and Neurotoxicity. Anesth Analg 2018;127:e16. doi: 10.1213/ANE.0000000000003474.
- (II) Hahn RG. Arterial Pressure and the Rate of Elimination of Crystalloid Fluid. Anesth Analg 2017;124:1824-33. doi: 10.1213/ANE.0000000000002075.
- (III) Katherine B. Hagan; Acceptance: A Letter to My Grieving Self. Anesthesiology 2018;129:1046-7. doi: 10.1097/ALN.0000000000002345.
- (IV) Viarasilpa T, Panyavachiraporn N, Marashi SM, et al. Prediction of Symptomatic Venous Thromboembolism in Critically Ill Patients: The ICU-Venous Thromboembolism Score. Crit Care Med. 2020;48:e470-9. doi: 10.1097/CCM.0000000000004306.
- (V) Nie Y, Li S, Yan T, et al. Propofol Attenuates Isoflurane-Induced Neurotoxicity and Cognitive Impairment in Fetal and Offspring Mice. Anesth Analg 2020;131:1616-25. doi: 10.1213/ANE.0000000000004955.

Appendix 2 (prompts and output Table 1)

Prompt 1: Limitations of ChatGpt

Output: While ChatGPT and similar language models are powerful tools, they do have some limitations. Here are a few notable ones:

- (I) Lack of real-world understanding: ChatGPT lacks genuine understanding of the world. It generates responses based on patterns and correlations in the training data rather than actual comprehension. This can result in incorrect or nonsensical answers.
- (II) Sensitivity to input phrasing: The model can be sensitive to slight rephrasing of questions or prompts. Small changes in wording may yield different responses, which can be inconsistent or frustrating for users.
- (III) Propensity for biased or controversial outputs: ChatGPT learns from vast amounts of internet text, which may include biased or offensive content. As a result, it may occasionally generate responses that are politically biased, offensive, or perpetuate stereotypes.
- (IV) Inability to verify information: ChatGPT does not have access to real-time information or the ability to browse the internet. It cannot fact-check or provide the most up-to-date information beyond its training knowledge, which has a cutoff date.
- (V) Tendency to generate plausible but incorrect information: While ChatGPT can generate coherent and contextually appropriate responses, it can also produce answers that sound plausible but are factually incorrect. Users should independently verify important information.
- (VI) Difficulty handling ambiguous queries: If a user's question or prompt is ambiguous or lacks necessary context, ChatGPT may struggle to provide a relevant or accurate response. It may make assumptions or seek clarifications that human users wouldn't require.

It's essential to exercise caution and critical thinking when using ChatGPT or any AI language model. It is not a substitute for professional advice, and its responses should be evaluated accordingly.

Prompt 2: The history of AI

Output: The history of AI dates back several decades,

marked by significant milestones and advancements. Here's a brief overview of key developments:

- (I) Early concepts (1940s–1950s): The foundations of AI were laid during this period. Researchers like Alan Turing and John von Neumann proposed theoretical frameworks for machine intelligence, including the famous Turing Test.
- (II) Dartmouth Conference (1956): The term “artificial intelligence” was coined, and the field of AI was officially established. The Dartmouth Conference brought together pioneers like John McCarthy, Marvin Minsky, and Claude Shannon to discuss AI research and its potential.
- (III) Symbolic AI and expert systems (1960s–1980s): Symbolic AI, also known as “good old-fashioned AI” (GOFAI), focused on rule-based systems and symbolic reasoning. Expert systems, which used domain-specific knowledge to solve problems, gained popularity during this period.
- (IV) Neural networks and ML (1950s–1970s): Researchers started exploring neural networks as a computational model for AI. The perceptron, an early form of artificial neural networks, was introduced. However, interest in neural networks declined until the 1980s.
- (V) AI winter (1980s–1990s): Progress in AI research faced significant challenges, resulting in reduced funding and a period known as the “AI winter”. Optimism waned as the initial hype surrounding AI didn't match practical results.
- (VI) Rise of ML (2000s–present): ML, particularly with the advent of big data and computational power, brought renewed interest in AI. Techniques like deep learning, reinforcement learning, and data-driven approaches led to breakthroughs in various domains.
- (VII) NLP and chatbots (2010s–present): NLP made significant progress, leading to advancements in speech recognition, language understanding, and machine translation. Chatbots and virtual assistants emerged, leveraging NLP to interact with users.
- (VIII) AI applications and integration (present): AI technologies have been integrated into numerous applications and industries. AI is employed in image recognition, autonomous vehicles, recommendation systems, healthcare diagnostics, finance, and many other domains.

It's important to note that AI is an evolving field, and breakthroughs continue to occur. The timeline above provides a high-level overview, and there are numerous other important contributions and subfields within AI.

Prompt 3: AI subtypes

Output: AI encompasses various subfields and approaches. Here are some of the commonly recognized subtypes of AI:

- (I) **Narrow/weak AI:** Narrow AI refers to AI systems designed to perform specific tasks or functions with a narrow focus. These systems excel in a particular area but lack general intelligence. Examples include voice assistants, image recognition algorithms, and recommendation systems.
- (II) **General/strong AI:** General AI aims to possess human-level intelligence and proficiency across a wide range of tasks. This form of AI would have the ability to understand, learn, and apply knowledge in a manner similar to human beings. Achieving strong AI remains a significant challenge and is an active area of research.
- (III) **ML:** ML involves training algorithms to learn patterns and make predictions or decisions without explicit programming. ML algorithms learn from data and improve their performance over time. Supervised learning, unsupervised learning, and reinforcement learning are common approaches within ML.
- (IV) **Deep learning:** Deep learning is a subset of ML that focuses on training deep neural networks with multiple layers. These networks can learn hierarchical representations of data and have achieved remarkable success in image recognition, NLP, and other complex tasks.
- (V) **NLP:** NLP is concerned with enabling computers to understand, interpret, and generate human language. It involves tasks such as speech recognition, language understanding, sentiment analysis, and machine translation.
- (VI) **Computer vision:** Computer vision deals with enabling machines to understand and interpret visual data, such as images and videos. It involves tasks like object detection, image recognition, image segmentation, and scene understanding.
- (VII) **Robotics:** Robotics combines AI, ML, and control systems to develop intelligent machines

that can interact with the physical world. Robotics aims to create machines capable of perception, decision-making, and physical manipulation.

- (VIII) **Expert systems:** Expert systems are AI systems that leverage knowledge and rules provided by human experts in specific domains. They use this knowledge to solve complex problems, provide recommendations, or make decisions.
- (IX) **Reinforcement learning:** Reinforcement learning involves training agents to learn through interaction with an environment. Agents receive feedback in the form of rewards or penalties based on their actions, allowing them to learn optimal strategies or policies.

These are just a few examples of AI subtypes, and the field of AI is broad and continuously evolving. Many subfields and approaches intersect and complement each other to advance the development and application of AI.

Prompt 4: Example of rule-based AI

Output: One example of rule-based AI is an expert system used in medical diagnosis. These systems utilize a set of predefined rules and knowledge provided by domain experts to assist in diagnosing diseases or medical conditions.

Here's a simplified example:

- (I) **Symptom input:** The user provides information about their symptoms, such as fever, cough, and headache.
- (II) **Rule-based inference:** The expert system uses a set of rules, typically in the form of "if-then" statements, to reason and make decisions. For instance:
 - ❖ If fever is present and cough is present, then it could be a respiratory infection;
 - ❖ If fever is present and headache is present, then it could be influenza;
 - ❖ If cough is present for more than 2 weeks, then it could be bronchitis.
- (III) **Decision and recommendation:** Based on the rules and the user's symptom input, the expert system generates a diagnosis or a list of possible conditions. It may also provide recommendations, such as seeking further medical evaluation or suggesting self-care measures.

Iterative process: The user's responses and additional

questions from the system can help narrow down the possible diagnoses or provide more specific recommendations. The expert system continues to refine its reasoning based on the provided information and the rule-based knowledge.

Rule-based AI systems like expert systems are designed to emulate the decision-making process of human experts in a specific domain. They rely on explicit rules and knowledge representation, making them transparent and understandable. However, they can be limited by the comprehensiveness and accuracy of the rules and knowledge base they utilize.

Prompt 5: ethical and social issues in AI

Output: The rapid development and deployment of AI technologies have raised various ethical and social concerns. Here are some key ethical and social issues associated with AI:

- (I) Bias and fairness: AI systems can inherit biases from the data they are trained on, leading to discriminatory outcomes. This can result in unfair treatment or perpetuation of social biases related to race, gender, or other protected attributes. Ensuring fairness and addressing biases in AI algorithms is crucial.
- (II) Privacy and security: AI technologies often rely on vast amounts of personal data. Concerns arise regarding the collection, storage, and use of this data, as well as the potential for unauthorized access, breaches, or misuse. Protecting privacy rights and ensuring robust security measures are essential.
- (III) Accountability and transparency: AI systems can make decisions that have significant impacts on individuals and society. It is crucial to establish mechanisms for accountability and transparency, allowing users and stakeholders to understand how decisions are made and to hold responsible parties accountable for AI-related outcomes.
- (IV) Explainability and interpretability: Some AI algorithms, such as deep learning models, can be opaque and difficult to interpret. Lack of explainability hampers understanding of how AI systems arrive at their decisions, limiting trust, and hindering their adoption in critical domains like healthcare and justice.
- (V) Employment disruption: The automation

potential of AI raises concerns about job displacement and changes in the workforce. While AI can enhance productivity, it can also lead to unemployment and exacerbate socioeconomic inequalities. Preparing for the impact on employment and ensuring a just transition is necessary.

- (VI) Autonomous systems and decision-making: The development of autonomous AI systems, such as self-driving cars or autonomous weapons, raises ethical dilemmas. Questions surrounding responsibility, liability, and the potential for unintended consequences require careful consideration and regulation.
- (VII) Data governance and ownership: The use of data in AI raises questions about ownership, control, and governance. Ensuring that data is collected, used, and shared in an ethical and responsible manner while respecting individual rights is crucial.
- (VIII) Impact on human relationships: The increasing integration of AI technologies in daily life can have social implications. Human interactions, social skills, and empathy may be affected, and the potential for AI to replace or disrupt human relationships raises ethical considerations.

Addressing these ethical and social issues requires interdisciplinary collaboration, involving experts from various fields, policymakers, industry stakeholders, and the public. Striking a balance between innovation and responsible development of AI is crucial to harness its benefits while mitigating its potential risks.

Prompt 6: Give references for the previous query

Output: I apologize for any confusion, but as an AI language model, I don't have direct access to external sources or the ability to provide specific references. The information I generate is based on a mixture of licensed data, data created by human trainers, and publicly available data. I should note that while I strive to provide accurate and up-to-date information, my responses may not always reflect the most current research or developments in the field of AI ethics.

If you require specific references, I recommend consulting academic databases like Google Scholar, IEEE Xplore, or ACM Digital Library. These platforms provide access to a wide range of scholarly articles and research papers on AI ethics and social implications. Additionally,

websites of reputable organizations dedicated to AI ethics and policy, such as the Partnership on AI, AI Now Institute, or Future of Life Institute, often publish reports and resources that discuss ethical and social considerations in AI.

By referring to these sources, you can explore the scholarly discussions and research on the ethical and social issues surrounding AI in greater detail.

Prompt examples

Certainly! Here are a few examples of references and resources related to the ethical and social issues in AI:

- (I) “Artificial Intelligence as Structural Estimation: Economic Interpretations of Deep Blue, Bonanza, and AlphaGo” by Susan Athey and Guido W. Imbens (link: <https://pubs.aeaweb.org/doi/10.1257/jep.31.2.237>);
- (II) “Fairness in Machine Learning: Lessons from Political Philosophy” by Solon Barocas and Andrew D. Selbst (link: <https://arxiv.org/abs/1712.03586>);
- (III) “The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation” by Brundage *et al.* (link: <https://arxiv.org/abs/1802.07228>);
- (IV) “Ethics of Artificial Intelligence and Robotics” by Vincent C. Müller (link: <https://plato.stanford.edu/entries/ethics-ai/>);
- (V) “The Alignment Problem: Machine Learning and Human Values” by Brian Christian (book);
- (VI) “Data and Society Research Institute” (link: <https://datasociety.net/>);
- (VII) “AI Now Institute” (link: <https://ainowinstitute.org/>);
- (VIII) “IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems” (link: <https://ethicsinaction.ieee.org/>).

These references cover a range of topics, including fairness in ML, the ethical implications of AI, and the social impact of AI. Please note that these references are for illustrative purposes, and it’s always advisable to conduct your own research and explore additional sources to gain a comprehensive understanding of the subject.

Prompt 7: AI in anesthesia

Output: AI is showing promising applications in the field of

anesthesia, offering potential improvements in patient care, efficiency, and decision support. Here are a few examples of how AI is being used in anesthesia:

- (I) Anesthetic drug delivery: AI algorithms can analyze patient data, including vital signs, drug concentrations, and response to anesthesia, to optimize drug dosing and delivery. These systems can provide real-time guidance to anesthesiologists, helping them maintain appropriate anesthesia levels and reduce the risk of complications.
- (II) Predictive analytics: AI models trained on large datasets can predict patient-specific risks and outcomes during anesthesia. By analyzing factors such as patient demographics, medical history, and preoperative assessments, AI algorithms can provide insights into the likelihood of complications, length of stay, or postoperative pain, aiding in preoperative planning and decision-making.
- (III) Automated monitoring: AI-based monitoring systems can continuously analyze patient vital signs, such as heart rate, blood pressure, oxygen saturation, and respiratory rate. These systems can detect abnormalities or trends that may indicate anesthesia-related complications, allowing early intervention and improved patient safety.
- (IV) Decision support systems: AI algorithms can assist anesthesiologists in making critical decisions during surgery. By considering patient-specific data, surgical parameters, and real-time monitoring information, AI systems can provide recommendations on anesthetic techniques, fluid management, or adjustments in drug dosages.
- (V) Risk stratification: AI models can analyze electronic health records and historical data to identify patients at higher risk for anesthesia-related complications. By flagging high-risk patients, anesthesiologists can provide tailored care and implement preventive measures to minimize adverse events.
- (VI) Education and training: AI-based simulation platforms offer virtual environments for anesthesiology training and education. These systems provide realistic scenarios and feedback, allowing trainees to practice anesthesia techniques, decision-making, and crisis management in a safe and controlled setting.

It's important to note that while AI has the potential to enhance anesthesia care, these technologies are still evolving, and their implementation should consider regulatory requirements, clinical validation, and ongoing human supervision. Anesthesiologists continue to play a crucial role in overseeing patient care and making final decisions based on their expertise and judgment.

Prompt 8: Give references for the previous query_examples

Output: Certainly! Here are a few examples of studies and articles that discuss the use of AI in anesthesia:

- (I) "Deep Learning in Anesthesia: Not Yet Ready for Prime Time?" by Pakdaman *et al.* (Link: <https://pubmed.ncbi.nlm.nih.gov/30550490/>);
- (II) "Application of Artificial Intelligence in Anesthesia and Intensive Care" by Cannesson *et al.* (Link: <https://pubmed.ncbi.nlm.nih.gov/29945526/>);
- (III) "Development and Validation of an Artificial Intelligence Model for the Prediction of Hypotension During Surgery" by Le *et al.* (Link: <https://pubmed.ncbi.nlm.nih.gov/31984360/>);
- (IV) "Artificial Intelligence in Anesthesia and Intensive Care" by Sebastian *et al.* (Link: <https://pubmed.ncbi.nlm.nih.gov/33594302/>);
- (V) "Artificial Intelligence and Machine Learning in Anesthesia" by Williams and Cannesson (Link: <https://pubmed.ncbi.nlm.nih.gov/31479430/>).

Prompt 9: Write a conclusion for the text

The term "artificial intelligence (AI)" was first used by John McCarthy, a computer scientist, in 1956 during a conference held in the summer of 1956 at Dartmouth College in Hanover, New Hampshire. It was a significant event in the history of AI. During the meeting, McCarthy, Marvin Minsky, Nathaniel Rochester, Claude Shannon, and the attendees used the term "artificial intelligence (AI)" to describe a field of study focused on creating machines that can exhibit intelligent behavior and perform tasks that would normally require human intelligence. The Dartmouth Conference is considered the birth of AI as a field of study, and it laid the foundations for the development of AI technology in the years that followed.

Another founder of the discipline is Alan Turing. He was a mathematician and computer scientist who is widely considered to be the father of theoretical computer science.

Six years before the conference, in 1950, Turing published an article in the journal *Mind* titled "Computing machinery and intelligence", in which he introduced the concept of the "Turing Test" as a way to determine whether a machine can exhibit human-like intelligent behavior that is indistinguishable from a human. In the article, Turing argued that if a machine can carry on a conversation with a human in such a way that the human is unable to distinguish it from another human, then it can be considered intelligent. Turing's ideas about AI and the Turing Test have had a significant impact on the field of computer science and have influenced the development of the AI discipline.

Beyond the historical notes, AI refers to the ability of a computer or machine to perform tasks that would normally require human intelligence, such as learning, problem-solving, and decision-making. Ultimately, the objective of AI is to develop systems capable of executing tasks that typically require human intelligence, with the aim of enhancing efficiency and precision.

AI subtypes

There are several different types of AI, including Rule-based AI, ML, Neural network, and Expert system AI.

Rule-based AI

This type of AI follows a set of predetermined rules to perform a task. It is based on the idea of creating a set of explicit rules that a machine can follow to solve a problem. These subtypes of AI are relatively simple and are often used to perform specific tasks where the rules are well-defined, and the problem domain is limited.

In medicine, a rule-based AI system designed to diagnose medical conditions might be programmed with a set of rules that describe the symptoms of different diseases. For example, the system might be programmed with the rule: "If a patient has a fever and a cough, then they may have the flu". The system would then use this rule, along with other rules about the symptoms of different diseases, to diagnose a patient's condition based on their symptoms.

ML

This type of AI uses algorithms to learn from data and improve its performance over time. It is based on the idea of creating algorithms that can automatically learn and improve from experience, without being explicitly programmed.

There are several different types of ML algorithms that can be classified into two main categories: supervised

and unsupervised. In supervised learning, the algorithm is trained on a dataset that has been labeled with the correct output or result. This allows the algorithm to learn how to map inputs to outputs, and it can then make predictions or decisions on new data based on its learning. Examples of supervised learning algorithms include linear regression, logistic regression, and support vector machines. Unsupervised learning involves the use of an algorithm that is not given any labeled data. Instead, the algorithm must find patterns and structure in the data on its own. This type of ML is often used for tasks such as clustering, in which the goal is to group similar items together. The most used unsupervised learning algorithms include k-means clustering and principal component analysis.

There are also other subtypes of ML algorithms, such as semi-supervised learning, reinforcement learning, and deep learning. These algorithms are typically used for more complex tasks and involve advanced techniques and approaches.

In medicine, ML is used for diagnosis, predictive modeling, drug discovery, and clinical decision support. For diagnostic aims, ML algorithms can be trained on medical data to recognize patterns and make accurate diagnoses. For example, a ML algorithm trained on medical images could be used to identify cancerous tumors. ML algorithms can be also used to develop predictive models. They can analyze patient data and make predictions about the likelihood of certain outcomes, such as the risk of developing a particular disease or the likelihood of a patient responding to a particular treatment. Drug discovery is an important application of ML. These strategies can be implemented to analyze chemical compounds and predict their potential as drugs, helping to speed up the drug discovery process. Finally, ML can provide recommendations and assist in decision-making for clinicians, helping to improve patient care and reduce errors.

Neural networks

This type of AI is inspired by the way the human brain works and is composed of interconnected “neurons” that process and transmit information. It is based on the idea of creating a network of artificial neurons that can process and transmit information in a way that is like the human brain.

Each neural network is composed of a variable number of interconnected processing nodes, or “neurons”, which are organized into layers. The input layer receives data, which is then processed and transmitted through one or more hidden layers before reaching the output layer. Each layer of neurons processes the data and passes it on to the next layer,

with the goal of recognizing patterns and relationships in the data. The connections between neurons are weighted, which means that each connection has a specific value that determines the strength of the connection. During the training process, the weights of the connections are adjusted to improve the accuracy of the neural network. Several types of neural networks have been developed. They can be broadly classified based on their architecture and the type of problem they are designed to solve (*Figure 1*). Some common types of neural networks include feedforward neural networks (the data flows in one direction from the input layer to the output layer without looping back) used for classification and prediction, CNNs to process data with a grid-like topology, such as images, RNNs to process sequential data, such as time series or natural language (e.g., for language translation), and autoencoders used for dimensionality reduction and feature learning (encoder and decoder elements work together to learn a compressed representation of the data). There are also many other types of neural networks and variations on these basic architectures, such as LSTM networks and GANs.

One of the key advantages of neural networks is their ability to learn and adapt to new data. As the neural network is exposed to more data, it can improve its performance and accuracy by adjusting the weights of the connections between neurons.

Neural networks are often used for tasks such as image and speech recognition, language translation, and forecasting. They are particularly well-suited for tasks that involve complex patterns and relationships in the data. However, they can also be computationally intensive and require large amounts of data for training.

Expert system AI

It is designed to mimic the decision-making abilities of a human expert in a particular field. Expert AI is based on the idea of creating a system that can use knowledge and reasoning to solve problems and make decisions in a way that is similar to a human expert. Thus, expert system AI is often implemented using a “knowledge base” that contains information about a particular domain, such as medical knowledge or technical knowledge. The system is then able to use this information to reason about problems and make decisions. These approaches have been used in a variety of fields, including medicine, finance, and customer service.

Some common algorithms used in expert system AI include rule-based algorithms (they use a set of explicit

rules), case-based reasoning algorithms (past examples or “cases” to solve new problems), decision tree algorithms (tree-like structure to represent the possible decisions and outcomes of a problem), and Bayesian networks (represent the uncertainties and dependencies of a problem).

Ethical and social issues in AI

Since AI is a rapidly developing field, many ethical and social issues need to be considered as it continues to advance, especially in fields such as medicine. For example, AI systems may be biased against certain groups of people, either intentionally or unintentionally. This can occur if the data used to train the AI system or the algorithms adopted to develop the AI system are unbalanced or inaccurate. Moreover, there are concerns about the potential for AI systems to infringe upon privacy, either by collecting and storing personal data or by using that data in ways that are not transparent or controlled by the individual. Another issue is AI autonomy. Notably, there are concerns about the increasing reliance on AI strategies and the potential for these systems to make decisions without human oversight. This raises questions about the accountability and responsibility of AI systems and their potential to infringe upon human autonomy. Other problems concern unemployment and transparency. Probably, the increasing use of AI may lead to job displacement and unemployment, as machines and algorithms may be able to perform certain tasks more efficiently than humans. Finally, a special issue is the “black box” nature of many AI systems as it can make it difficult to understand how they make decisions and to hold them accountable for their actions. XAI is a field of AI that focuses on developing systems that can provide explanations for their actions, predictions, or decisions. It can help to build trust in AI systems, especially when they are used in sensitive applications such as healthcare, finance, or criminal justice. Consequently, there is a need for appropriate regulation of AI to ensure that it is developed and used ethically and responsibly.

AI in anesthesia

Remarkably, AI has the potential to revolutionize the field of anesthesiology by improving patient care and safety. There are several ways in which AI can be used in anesthesiology. For example, for guiding anesthetic drug dosing, perioperative monitoring, and predictive modeling, as well as education and training.

Anesthetic drug dosing

Anesthetic drug dosing is a complex process that requires careful consideration of a number of factors, including a patient’s age, weight, medical history, and the type of surgery being performed. Misjudging the dosage can have serious consequences, including inadequate anesthesia, which can result in patient discomfort and increased risk of complications, or oversedation, which can lead to potentially fatal outcomes.

AI has the potential to assist in the precise dosing of anesthetic drugs by using algorithms to analyze patient data and provide recommendations for dosage. A number of studies have demonstrated the effectiveness of AI in this regard. For example, it was demonstrated that an AI-based system was able to accurately predict the optimal propofol dosage for individual patients during surgery, leading to improved patient outcomes and reduced risk of adverse events. In another investigation, the authors developed and validated an ML model based on a decision tree algorithm for predicting the requirement of propofol during surgery. The model was trained on data from a large number of patients and was able to accurately predict the optimal propofol dosage for individual patients. Intraoperative opioid administration was also investigated, and authors developed and validated a model based on a random forest algorithm for predicting the intraoperative requirements of propofol and remifentanyl.

Monitoring

Another potential application is monitoring. AI-powered monitoring systems can continuously analyze patient data and alert anesthesiologists to potential problems. These systems can also provide recommendations for treatment and can help to reduce the risk of complications.

There have been a number of studies that have explored the use of AI for anesthesia monitoring. For example, some researchers have developed ML algorithms that can analyze data from various anesthesia monitors (e.g., electrocardiogram, blood pressure, etc.) to identify patterns that may indicate a change in a patient’s condition. Other studies have focused on using NLP to automatically extract information from electronic medical records and other sources to aid in decision-making during anesthesia.

Predictive models for outcomes

AI can also be employed for the development of predictive models: AI algorithms can analyze patient data and predict the likelihood of certain outcomes, such as the need for

additional anesthetic drugs or the risk of postoperative complications. This can help anesthesiologists to make more informed decisions about patient care.

One study used ML algorithms to develop predictive models for PONV based on patient characteristics and anesthetic factors. The model was able to accurately predict the likelihood of PONV in patients, allowing anesthesiologists to tailor their treatment accordingly. Other investigations focused on predictive modeling for postoperative pain management. For example, in a retrospective analysis of 2,200 patients, it was found that an ML algorithm was able to predict the likelihood of patients experiencing moderate to severe pain after surgery based on patient characteristics and surgical factors. According to the authors, the predictive model developed in the study could potentially be used to proactively manage pain and improve patient outcomes.

In a systematic review and meta-analysis, the authors analyzed the available evidence on the use of AI in anesthesia. The review found that AI has the potential to improve patient safety and outcomes, but also highlighted the need for further research to assess the effectiveness and safety of AI in this field. In a more recent evidence-based analysis, the authors investigated different postoperative outcomes, including mortality risk, systemic complications (e.g., pulmonary, cardiovascular, acute kidney injury), intensive care unit admission, anesthesiologic risk, and prolonged length of hospital stay. They reported that the most frequently used algorithms were gradient boosting, random forest, logistic regression, artificial neural networks, and support vector machines. The models with the best performance were random forest and gradient boosting (AUC >0.90). The authors found that using ML algorithms can result in more accurate prediction of outcomes compared to validated prognostic scores and traditional statistical methods.

Education and training

Applications of AI can be used to create virtual simulations and training programs that can help anesthesiologists to improve their skills and knowledge. Different studies demonstrated the potential of AI to assist in the assessment of trainee performance in anesthesia. For instance, it was developed and validated a ML model for automated assessment of trainee anesthetic performance. The model was trained on data from a large number of trainees and was able to accurately predict the performance of individual participants based on their recorded data. Furthermore, in

their systematic review, the authors analyzed the available evidence on the use of AI in anesthesia education and proved that AI has the potential to improve the efficiency and effectiveness of training and assessment in anesthesia, but also highlighted the need for further research to assess the effectiveness and safety of AI in this context.

Output: in conclusion, the term “artificial intelligence (AI)” originated in 1956 during the Dartmouth Conference, where pioneers like John McCarthy introduced the concept of creating machines capable of exhibiting intelligent behavior. Alan Turing’s contributions, particularly his concept of the Turing Test, further shaped the field of AI and influenced its development. AI aims to enable computers and machines to perform tasks that typically require human intelligence, with the goal of enhancing efficiency and precision.

AI encompasses various subtypes, including rule-based AI, ML, neural networks, and expert systems. Rule-based AI follows predetermined rules to solve problems within well-defined domains. ML algorithms enable machines to learn from data and improve their performance over time, either through supervised or unsupervised learning. Neural networks, inspired by the human brain, process and transmit information through interconnected artificial neurons, making them suitable for tasks involving complex patterns. Expert systems mimic human decision-making abilities using knowledge and reasoning from a specific domain.

However, the rapid advancement of AI raises ethical and social concerns. Biases in AI systems, privacy infringement, lack of human oversight, unemployment, and the opacity of decision-making processes are some of the challenges that need to be addressed. The development and use of AI must be regulated to ensure ethical and responsible practices, with XAI being an emerging field that focuses on developing transparent AI systems.

In the field of anesthesiology, AI holds great promise for improving patient care and safety. AI can assist in precise anesthetic drug dosing, continuous monitoring of patient data, predictive modeling for outcomes, and education and training of anesthesiologists. By leveraging AI algorithms, anesthesiologists can enhance dosing accuracy, detect potential problems, predict patient outcomes, and improve training programs.

While AI has shown potential in anesthesia, further research is needed to assess its effectiveness and safety. Nevertheless, the integration of AI in anesthesiology has the potential to revolutionize the field, leading to improved patient outcomes and advancing the practice of anesthesia.

Appendix 3 (checklist for text evaluation Table 1)

Prompt: Limitations of ChatGpt
Options:
Relevant and clearly expressed output;
Too succinct and not very detailed;
Not relevant.

Prompt: The history of AI
Options:
Relevant and clearly expressed output;
Too succinct and not very detailed;
Not relevant.

Prompt: AI subtypes
Options:
Relevant and clearly expressed output;
Too succinct and not very detailed;
Very useful information.
Not useful information

Prompt: Example of Rule-based AI
Options:
Pertinent output;
Not pertinent.

Prompt: Ethical and social issues in AI
Options:
Pertinent output;
Generic response;
Very useful information;
Not useful information.

Prompt: Give references for the previous query
Options:
Correct references;
Incorrect references;
Very useful information;
Not useful information.

Prompt: AI in anesthesia
Options:
Complete list;
Not complete list;
Useful information;
Partially useful information;
Not useful information.

Prompt: Give references for the previous query
Options:
Correct references;
Incorrect references;
Very useful information;
Not useful information.

Prompt: Write a conclusion from the text
Options:
Exhaustive summary;
Brief summary;
Not useful text.