Brief Communication

Artificial intelligence hallucinations in anaesthesia: Causes, consequences and countermeasures

Artificial intelligence (AI) hallucinations occur when large language models, such as chatbots or computer vision systems, generate outputs containing non-existent patterns, leading to inaccurate results. Also known as AI confabulations or delusions, these instances challenge expectations of appropriate responses from AI tools due to unrelated or pattern-lacking outputs, similar to human hallucinations. Addressing such issues with generative AI presents significant challenges despite ongoing efforts to resolve them.[1,2]

CAUSES OF AI HALLUCINATIONS

Various causes of AI hallucinations have been identified and include:

Insufficient or biased training data: An AI model designed to assist anaesthesiologists in administering anaesthesia may be trained predominantly on data from patients of a certain demographic, such as adults of average weight. When faced with a paediatric patient or an obese patient, the AI model may possibly hallucinate dosage recommendations that are inaccurate or unsafe, as it lacks sufficient exposure to diverse patient populations.[3]

Model complexity: A highly complex AI system tasked with monitoring vital signs during surgery may exhibit hallucinatory responses when encountering unusual physiological patterns. This complexity underscores the need for simpler models to avoid such hallucinations.[4]

Lack of explainability (black box): An AI algorithm designed to predict anaesthesia induction times may produce unexpectedly long or short estimates without providing clear explanations for its predictions. In cases where anaesthesiologists cannot understand or verify the AI system's reasoning, there is a risk of blindly following its recommendations, potentially leading to errors or patient harm. This highlights the urgent need for explainable AI in anaesthesia.[5]

MULTIFACETED THREAT OF AI HALLUCINATIONS IN ANAESTHESIA

An AI hallucination occurs when an AI system produces demonstrably incorrect or misleading outputs, appearing confident and plausible despite factually flawed. The possible impacts of AI hallucinations on anaesthesia domains are varied[6-9] [Table 1].

Misdiagnosis and mistreatment: Hallucinations can misinterpret patient data, resulting in unnecessary interventions or delayed treatments.

Medication errors: AI-driven systems may recommend incorrect drug dosages, impacting patient safety.

Communication and documentation: Misinterpreted verbal commands or procedure details can hinder accurate documentation and patient safety.

Research skewing: AI-driven analysis of anaesthesia data for research could be skewed by hallucinations, leading to misleading conclusions.

Legal and ethical concerns*:*

Liability: Who is responsible for the errors caused by AI hallucinations? This remains a complex question with no clear answer. Depending on the specific circumstances, potential targets include the AI developer, healthcare provider or hospital.

Informed consent: How can patients be adequately informed about the risks of AI hallucinations in anaesthesia, given the technical complexity involved and the dynamic nature of AI outputs? Striking a balance between transparency and patient anxiety is crucial.

Bias: AI algorithms can perpetuate societal biases, leading to discriminatory outcomes in health care. Imagine an AI system trained on biased data; it might recommend different treatments based on a patient's race or socioeconomic background.[10-12]

STRATEGIES TO MITIGATE AI HALLUCINATIONS

Various mitigation strategies need to be adhered to for the impact of AI hallucination on health care [Figure 1].

High-quality, diverse training data: Utilising diverse datasets improves AI model accuracy and reduces hallucination risks. For example, research by Jones

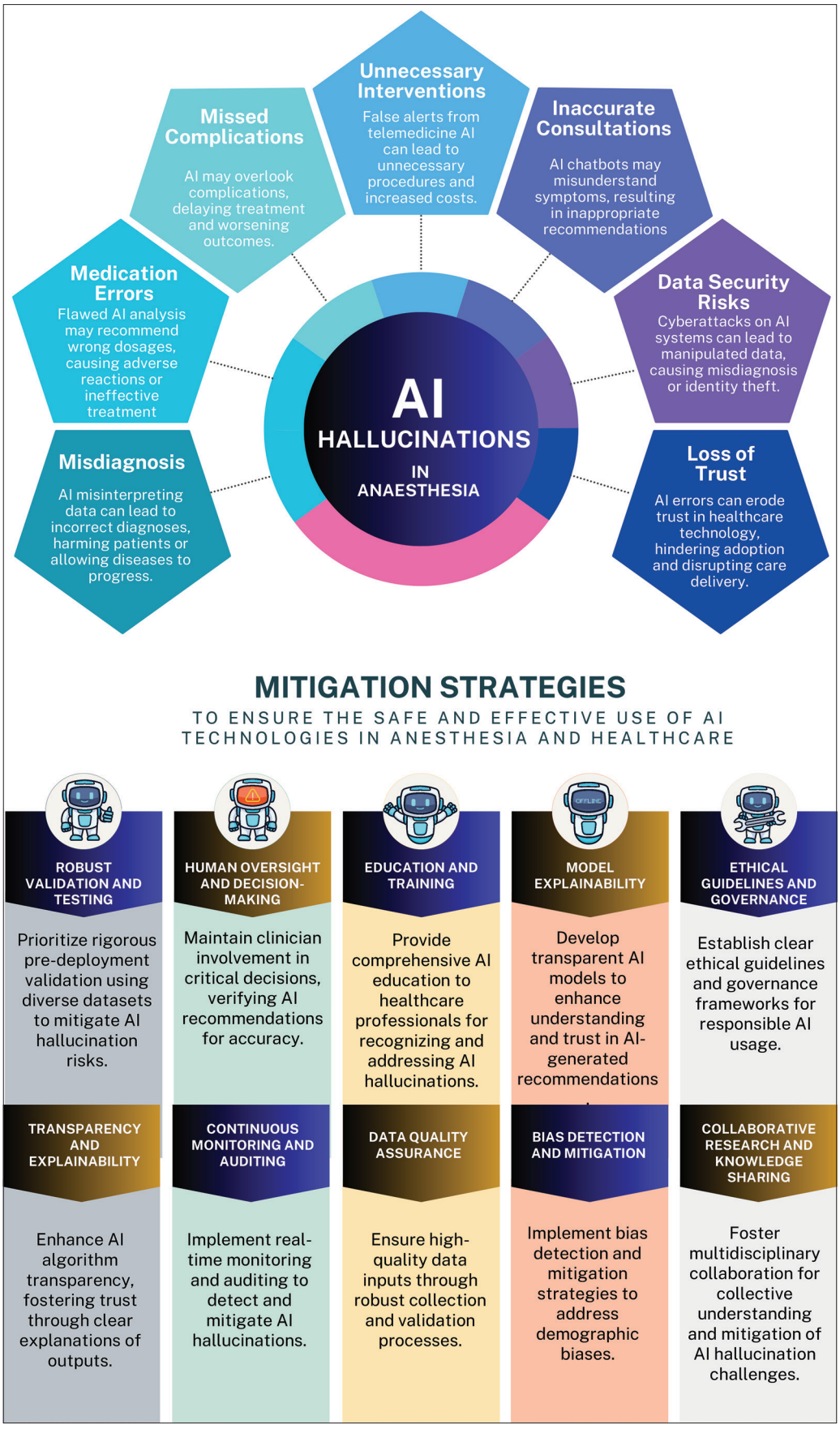

Figure 1: Impact of AI hallucination on health care and mitigation strategies. AI = artificial intelligence

Table 1: Examples of AI hallucinations' possible impact on anaesthesia domains

Anaesthesia domain	AI impact on anaesthesia	AI hallucination concerns
Pre-anaesthesia evaluation	AI optimises pre-anaesthesia assessments by parsing patient data, identifying risk factors and offering tailored recommendations	The emergence of hallucinated data may introduce inaccuracies, potentially disrupting anaesthesia plans and patient management strategies
Preoperative phase	AI facilitates preoperative planning, outcome prognostication and interdisciplinary communication within healthcare teams	The risk of generating false prognostications or recommendations impacts surgical strategies and patient outcomes
Post-anaesthesia care unit	AI enhances post-anaesthesia care through continuous monitoring, complication prediction and timely intervention facilitation	Hallucinated data may trigger erroneous alerts or predictions, posing risks to effective patient management and postoperative care
Intraoperative management	AI fortifies intraoperative safety by augmenting monitoring, anaesthesia titration and surgical coordination	The generation of misleading monitoring data or alerts jeopardising medication dosages and surgical interventions
Patient communication	AI-driven tools streamline patient education and communication with healthcare providers, ensuring personalised interactions	AI hallucinations in communication platforms may propagate inaccurate or confusing information to patients, compromising the efficacy of healthcare delivery
Critical care medicine	AI enables early detection of patient deterioration, prognostication of outcomes and optimisation of treatment modalities	AI hallucinations within critical care monitoring systems may trigger false alarms or erroneous clinical predictions, posing risks to patient welfare
Pain medicine	AI aids in pain assessment, tailoring treatment plans and forecasting treatment responses	Inaccuracies in pain assessment or treatment predictions, thereby impacting patient care outcomes
Research	AI accelerates research endeavours by analysing datasets and expediting the development of novel therapeutic interventions	May introduce biases or inaccuracies, undermining the integrity of scientific investigations
Resident training programmes	AI-powered simulations enhance resident training by providing immersive experiences and refining clinical skills	AI hallucinations within simulation scenarios may create unrealistic or hazardous training environments, impeding resident education and skill development
Education	AI-driven educational platforms offer trainees personalised learning experiences and real-time feedback, fostering skill acquisition	AI hallucinations in educational materials may impart misleading information or feedback to learners, hindering the effectiveness of educational interventions

AI=artificial intelligence

et al.[13] demonstrated how incorporating various demographic factors and medical histories in training data significantly improved the accuracy of an AI-driven diagnostic tool for skin cancer detection.

Explainable AI: Developing transparent AI models aids in identifying and rectifying hallucinations. For instance, the explainable nature of a deep learning model used in financial fraud detection allowed analysts to trace back erroneous predictions to specific data points, enabling targeted adjustments to the model's training data and architecture.[14]

Human oversight and collaboration: Human involvement reduces hallucination risks, especially in sensitive domains like health care. Collaborative efforts between AI systems and human experts have effectively reduced hallucination risks.[15]

Continuous monitoring and evaluation: Regular evaluation detects and addresses hallucinations promptly. Continuous monitoring of its AI-powered recommendation system and real-time user feedback analysis allows for swift identification and correction of hallucinated product suggestions, improving user satisfaction and trust.[16]

Algorithmic auditing and regulatory frameworks: Establishing robust auditing mechanisms and regulatory frameworks ensures AI system's accountability and reliability.[17]

To conclude, AI hallucinations in anaesthesia pose risks of misdiagnosis, medication errors and skewed research outcomes. Prioritising diverse training data, embracing explainable AI, maintaining human oversight, continuous monitoring and regulatory frameworks are crucial in mitigating these risks and fostering trust in AI technologies in health care.

Financial support and sponsorship
Nil.

Conflicts of interest
There are no conflicts of interest.

ORCID
Prakash Gondode: https://orcid.org/0000-0003-1014-8407

Sakshi Duggal: https://orcid.org/0000-0002-8865-0854
Vaishali Mahor: https://orcid.org/0009-0003-1251-122X

Prakash Gondode, Sakshi Duggal, Vaishali Mahor
Department of Anaesthesiology, Pain Medicine and Critical Care, All India Institute of Medical Sciences, New Delhi, India

Address for correspondence:
Dr. Prakash Gondode,
Department of Anesthesiology, Pain Medicine and Critical Care, All India Institute of Medical Sciences, New Delhi, India.
E-mail: drprakash777@gmail.com

Submitted: 24-Feb-2024
Revised: 11-Apr-2024
Accepted: 13-Apr-2024
Published: 07-Jun-2024

REFERENCES

1. Salvagno M, Taccone FS, Gerli AG. Artificial intelligence hallucinations. Crit Care 2023;27:180. doi: 10.1186/s13054-023-04473-y.
2. IBM. Think. "AI Hallucinations.". Available from: https://www.ibm.com/topics/ai-hallucinations. [Last accessed on 2024 Mar 16].
3. Ji Z, Lee N, Frieske R, Yu T, Su D, Xu Y, *et al*. Survey of hallucination in natural language generation. ACM Comput Surv 2023;55:1-38.
4. Hanneke S, Kalai AT, Kamath G, Tzamos C. Actively avoiding nonsense in generative models. In Conference on Learning Theory, 2018. PMLR. p. 209-27.
5. SCADS. Cracking the Code: The Black Box Problem of AI. Available from: https://scads.ai/cracking-the-code-the-black-box-problem-of-ai/#:~:text=The%20black%20box%20problem%20refers,This%20poses%20a%20significant%20challenge. [Last accessed on 2024 Mar 16].
6. Panch T, Mattie H, Atun R. Artificial intelligence and algorithmic bias: Implications for health systems. J Glob Health 2019;9:010318. doi: 10.7189/jogh. 09.020318.
7. Kumar M, Mani UA, Tripathi P, Saalim M, Roy S. Artificial hallucinations by Google Bard: Think before you leap. Cureus 2023;15:e43313. doi: 10.7759/cureus. 43313.
8. Chan KS, Zary N. Applications and challenges of implementing artificial intelligence in medical education: Integrative review. JMIR Med Educ 2019;5:e13930. doi: 10.2196/13930.
9. Morley J, DeVito NJ, Zhang J. Generative AI for medical research. BMJ 2023;382:1551. doi: 10.1136/bmj.p1551.
10. Hashimoto DA, Witkowski E, Gao L, Meireles O, Rosman G. Artificial intelligence in anesthesiology: Current techniques, clinical applications, and limitations. Anesthesiology 2020;132:379-94.
11. Harvey HB, Gowda V. Regulatory issues and challenges to artificial intelligence adoption. Radiol Clin North Am 2021;59:1075-83.
12. Keskinbora KH. Medical ethics considerations on artificial intelligence. J Clin Neurosci 2019;64:277-82.
13. Jones OT, Calanzani N, Saji S, Duffy SW, Emery J, Hamilton W, *et al*. Artificial intelligence techniques that may be applied to primary care data to facilitate earlier diagnosis of cancer: Systematic review. J Med Internet Res 2021;23:e23483. doi: 10.2196/23483.
14. Hassija V, Chamola V, Mahapatra A, Singal A, Goel D, Huang K, *et al*. Interpreting black-box models: A review on explainable artificial intelligence. Cogn Comput 2024;16:45-74.
15. Ueda D, Kakinuma T, Fujita S, Kamagata K, Fushimi Y, Ito R, *et al*. Fairness of artificial intelligence in healthcare: Review and recommendations. Jpn J Radiol 2024;42:3-15.
16. Necula SC, Păvăloaia VD. AI-driven recommendations: A systematic review of the state of the art in E-commerce. Appl Sci 2023;13:5531. doi: 10.3390/app13095531.
17. Mökander J. Auditing of AI: Legal, ethical and technical approaches. Digit Soc 2023;2:49. doi: 10.1007/s44206-023-00074-y.

Access this article online	
Quick response code	Website: https://journals.lww.com/ijaweb
	DOI: 10.4103/ija.ija_203_24

How to cite this article: Gondode P, Duggal S, Mahor V. Artificial intelligence hallucinations in anaesthesia: Causes, consequences and countermeasures. Indian J Anaesth 2024;68:658-61.