Exploring the benefits and challenges of AI-driven large language models in gastroenterology: Think out of the box

Jan Kral[1,2], Michal Hradis[3,4], Marek Buzga[5,6], Lumir Kunovsky[7,8,9]

Artificial Intelligence (AI) has evolved significantly over the past decades, from its early concepts in the 1950s to the present era of deep learning and natural language processing. Advanced large language models (LLMs), such as Chatbot Generative Pre-Trained Transformer (ChatGPT) is trained to generate human-like text responses. This technology has the potential to revolutionize various aspects of gastroenterology, including diagnosis, treatment, education, and decision-making support.

The benefits of using LLMs in gastroenterology could include accelerating diagnosis and treatment, providing personalized care, enhancing education and training, assisting in decision-making, and improving communication with patients. However, drawbacks and challenges such as limited AI capability, training on possibly biased data, data errors, security and privacy concerns, and implementation costs must be addressed to ensure the responsible and effective use of this technology.

The future of LLMs in gastroenterology relies on the ability to process and analyse large amounts of data, identify patterns, and summarize information and thus assist physicians in creating personalized treatment plans. As AI advances, LLMs will become more accurate and efficient, allowing for faster diagnosis and treatment of gastroenterological conditions.

Ensuring effective collaboration between AI developers, healthcare professionals, and regulatory bodies is essential for the responsible and effective use of this technology. By finding the right balance between AI and human expertise and addressing the limitations and risks associated with its use, LLMs can play an increasingly significant role in gastroenterology, contributing to better patient care and supporting doctors in their work.

Key words: artificial intelligence, large language model, gastroenterology

Received: April 14, 2024; Revised: July 21, 2024; Accepted: August 16, 2024; Available online: September 4, 2024
https://doi.org/10.5507/bp.2024.027

[1]Department of Internal Medicine, University Hospital Motol and Second Faculty of Medicine, Charles University, Prague, Czech Republic
[2]Department of Hepatogastroenterology, Institute for Clinical and Experimental Medicine, Prague, Czech Republic
[3]MAIA LABS s.r.o., Brno, Czech Republic
[4]Faculty of Information Technology, University of Technology, Brno, Czech Republic
[5]Department of Physiology and Pathophysiology, Faculty of Medicine, University of Ostrava, Ostrava, Czech Republic
[6]Institute of Laboratory Medicine, University Hospital Ostrava, Ostrava, Czech Republic
[7]2nd Department of Internal Medicine – Gastroenterology and Geriatrics, University Hospital Olomouc and Faculty of Medicine and Dentistry, Palacky University Olomouc, Olomouc, Czech Republic
[8]Department of Surgery, University Hospital Brno and Faculty of Medicine, Masaryk University, Brno, Czech Republic
[9]Department of Gastroenterology and Digestive Endoscopy, Masaryk Memorial Cancer Institute, Brno, Czech Republic
Corresponding author: Jan Kral, e-mail: jan.kral@centrum.cz

INTRODUCTION

The history of Artificial Intelligence (AI) as a discipline spans several decades and is marked by significant milestones. It all started in the 1950s when key concepts and early computing technologies were developed. In the 1960s and 1970s, researchers focused on symbolic AI and problem-solving techniques. However, the ambitious goals of the time led to the "AI winter" in the 1970s and 1980s. The 1980s and 1990s witnessed a resurgence of AI, with expert systems and different machine learning gaining prominence. Expert systems mimicked human expertise in specific domains, while neural networks around 2010 showed promise in pattern recognition and speech processing[1,2].

The turn of the century brought a paradigm shift in AI with the advent of machine learning, particularly deep learning. With increased computational power and access to vast amounts of data, deep learning algorithms revolutionized image recognition, natural language processing, and speech recognition[3]. Haug et al. in their review, of advancements in data science have allowed us to identify and analyse relationships within unstructured data, providing actionable insights into human behaviour. Neural networks, natural language processing and deep-learning models have made large language models (LLM) human-like interactions possible. This connectedness facilitated by data science has led to new discoveries in various fields, including social network analysis for counterterrorism, mining transactional data for business opportunities, and building grids of connected data for scientific research[4].

Today, AI has become an integral part of our lives, powering technologies like voice assistants, recommendation systems, and autonomous vehicles. Ethical considerations have gained prominence, prompting discussions on privacy, bias, and responsible AI development. Looking ahead, AI research focuses on advanced techniques such as reinforcement learning, generative models, and explainable AI (ref.[5]).

Current AI methods utilize algorithms and advanced computational techniques to mimic human intelligence and perform tasks traditionally requiring human intelligence. The methods involve processing large amounts of data, identifying patterns, and learning from examples. Machine learning algorithms enable AI systems to automatically improve their performance through exposure to data. Deep learning, a subset of machine learning, employs neural networks with multiple layers to extract complex features and make predictions. AI systems can employ various subdisciplines such as natural language processing, computer vision, and reinforcement learning. These systems continue to learn and adapt over time, becoming more proficient and accurate in their task performance[3,6].

Large Language Models

A Large Language Model (LLM) is a type of AI model designed to predict and generate text single word (token) at a time. These models are trained on vast amounts of text data which is approaching in its scope all the knowledge and experience humanity has collectively produced and shared online. To predict text as accurately as possible, these models have no choice but to understand text on ever deeper level, starting from lexical structures and syntax and ending with semantics at the level of human understanding. Contemporary LLMs, due to their size, can efficiently compress semantic concepts and memorize most of the information from their training data. Just by learning to predict text, large enough language models begin to display interesting and useful emergent capabilities such as text translation, summarization and document-based question answering[7]. These behaviours can be reliably elicited by suitable text prompts and further reinforced by either by supervised or reinforcement fine-tuning as is done with chat bots (e.g. ChatGPT) (ref.[8]).

The sizes of contemporary practical LLMs start around 7 billion parameters and end in the region of 500 billion parameters[8]. The small models can be deployed on a single consumer-grade PC with little effort. Large models require specialized hardware (from computation, memory, to high-bandwidth network interconnection) and software infrastructure for efficient operation. Training even the smallest LLMs requires multiple high-end processing units (GPU/TPU) and usually multiple distributed computational nodes representing a significant financial expense[9]. Training the larger models is a huge financial investment and engineering challenge to the point that computational clusters are designed and build specifically for the purpose of training LLMs and even the hardware is designed solely for this purpose.

Training basic LLMs from scratch is done mostly by companies and organizations with significant resources, fortunately some of them release such models to public under a range of licenses, some limited to research, some limited to non-commercial use and some without any restrictions. Such models can be fine-tuned for a specific purpose with relative ease and low cost by most organizations and even by individuals. In general, LLM fine-tuning should be understood as a way how to "explain" to the model what kind of output it is expected to produce, not as a process which should teach the model new knowledge or completely new capabilities.

Numerous LLMs have been developed by various organizations, some tailored for distinct applications and purposes. Examples of influential and widely used models are:

- ChatGPT (OpenAI): State-of-the-art proprietary conversational models by OpenAI.
- Gemini: Family of state-of-the-art proprietary conversational and multi-modal models developed by Google DeepMind.
- Mixtral-8x7B: State-of-the-art LLM with unrestrictive Apache 2.0 license.
- Llama 2: Models with 7, 13 and 70 billion parameters released under permissive license by Meta.
- MPT: Family of models with commercial license by MosaicML.
- Falcon: Family of models up to 180 billion parameters with commercial license by Technology Innovation Institute.
- BERT (Google): Context understanding in search queries.
- T5 (Google): Converts Natural Language Processing (NLP) tasks to text-to-text.
- Bard (Google): Google's conversational AI.
- LaMDA (Google): For open-ended conversations.
- XLNet (Google/CMU): Transformer model with permutation training.
- ERNIE (Baidu): Integrates knowledge graphs in language learning.
- RoBERTa (Facebook): Optimized BERT variant.
- ELECTRA (Google): Distinguishes real and fake input tokens.
- DeBERTa (Microsoft): Enhances BERT with disentangled attention.
- CTRL (Salesforce): For controllable text generation.
- BlenderBot (Facebook): Blends various conversational skills.
- Megatron (NVIDIA): For large-scale training.
- Jurassic-1 (AI21 Labs): Nuanced, context-aware responses.

Training a conversational LLM is an intricate and prolonged process, involving a series of detailed and precise and thorough steps: **Step 0**: The initial step in training a LLM involves the fundamental training of a general language model. **Step 1:** In supervised fine-tuning, the model learns to generate useful responses on a dataset of conversations written by humans. **Step 2:** A Reward

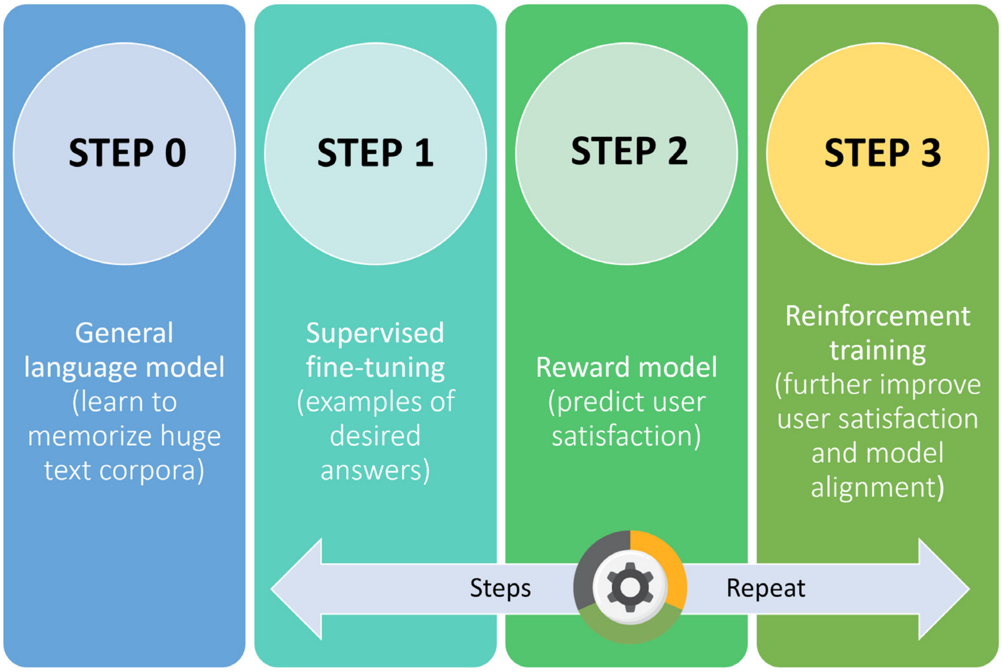

Fig. 1. ChatGPT training process (adapted and improved from Scalablepath.com)[10].

Model is trained using the outputs of the Supervised Fine-tuning Model. Human annotators rank the model's predictions based on their usefulness. This data is used to train the Reward Model to predict the quality of responses. **Step 3:** The Reinforcement Learning process is employed to further train the fine-tuned Model. The model acts as an agent, generating responses to user prompts. The responses are evaluated by the Reward Model, and the model updates its predictions to maximize the reward. This process is repeated to improve the model's performance (Fig. 1) (ref.[10]).

Latest LLMs approach or even exceed human performance in many tasks including math and legal exams, coding challenges and text comprehension[11,12]. It remains uncertain how LLMs could serve as a collaborative tool, potentially aiding researchers in improving their writing in the field of medicine itself[13-16]. LLMs can assist users in various tasks, such as answering questions, providing explanations, giving suggestions, and engaging in conversation on a wide range of topics. They have the ability to understand and generate text in multiple languages. Users interact with LLMs by providing prompts or questions, and they respond with text-based answers or suggestions.

While LLMs are highly skilled at generating human-like responses, it's important to note that they may occasionally produce incorrect or nonsensical information. Therefore, it is advisable to critically evaluate the information provided by LLMs and verify it from reliable sources when accuracy is crucial[17].

Potential benefits of using LLMs is gastroenterology (Fig. 2)

- **Accelerating diagnosis and treatment:** LLMs can swiftly enhance diagnosis and treatment for gastrointestinal issues by analyzing vast datasets of medical records. Trained on extensive patient information, they identify patterns unseen by doctors, leading to faster, more accurate diagnoses and tailored treatment plans. For example, in obesity management, this means creating personalized treatment plans that encompass dietary guidelines and behavioral modifications. Enhancing care for patients with colorectal cancer involves not only administering treatment but also proactively managing potential chemotherapy side effects and identifying possible medication interactions.
- **Personalized treatment:** Each patient is unique and requires an individualized approach. LLMs can analyse patient data and propose treatments tailored to their needs, improving the quality of care, and increasing the chances of successful treatment. LLMs can serve as an invaluable assistant or virtual colleague for clinicians. LLMs can recommend suitable biologic therapies based on a patient's current medications, past disease progression, and history of biological and surgical treatments for inflammatory bowel disease (IBD).
- **Results interpretation:** LLMs can interpret endoscopy results, pathology reports, and imaging scans like CT, MRI, and ultrasounds. They can also recommend subsequent diagnostic steps, treatment options, follow-up schedules, and highlight potential risks for patients, aiding in more informed and precise medical decision-making.
- **Education and training:** LLMs can be a valuable tool for educating and training doctors, medical students, and healthcare staff. They can provide information on the latest research articles, studies, and procedures in gastroenterology, making it easier to keep up with

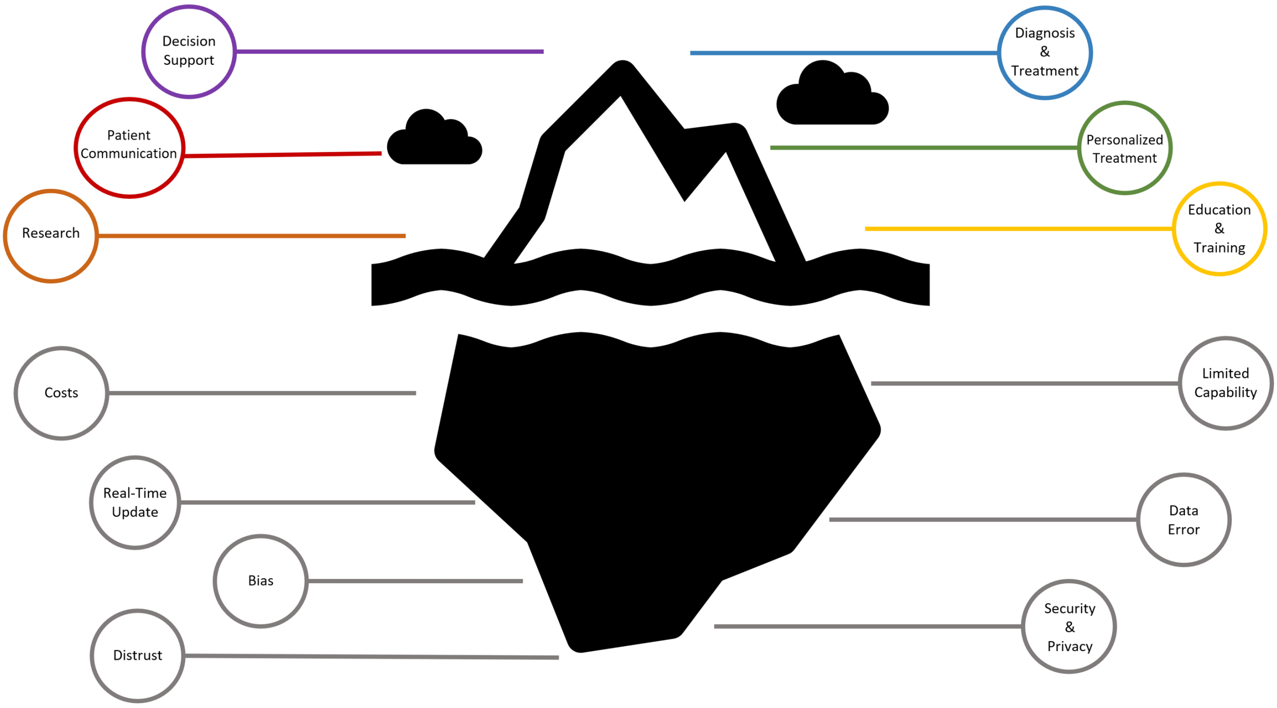

Fig. 2. Cons and Prons.

the ever-evolving field. With interactive scenarios and simulations, LLMs can offer hands-on experience in diagnosing and treating gastrointestinal diseases.

- **Research:** In gastroenterology research, LLMs can play a pivotal role by detecting trends and patterns in vast amounts of clinical data. This detection capability facilitates the identification of potential areas for advancement and steers the direction of new studies (especially in IBD, microbiome and cancer research). Consequently, it paves the way for the innovation of patient care techniques and the establishment of best practices within the field. Finally, AI and LLMs could assist with statistical analysis of data in studies and help researchers understand statistical methodologies and results.
- **Improve communication:** LLMs can bridge communication between doctors and patients via chat apps, enabling easy queries about conditions and treatments. They help simplify diagnoses and anticipate patient questions, improving understanding and consultation quality. Moreover, LLMs can help overcome language barriers when treating foreign patients.

Drawbacks of using LLMs in gastroenterology (Fig. 2)

- **Limited AI capability:** Although AI, like LLMs, offers many advantages, it cannot fully replace human judgment and experience (at the moment). AI can provide useful suggestions, but the final decision on diagnosis and treatment should be left to the experts.
- **Data errors:** ChatGPT relies on the data it analyses. If the input data is inaccurate or incomplete, this can lead to incorrect conclusions and inappropriate treatment recommendations.
- **Security and privacy concerns:** Sharing sensitive health information with AI can pose risks to patients' privacy. It is essential to ensure that data is secured and protected from unauthorized access.
- **Implementation costs:** Integrating AI, like ChatGPT, into gastroenterology practice can be costly. The expenses for technology, staff training, and system maintenance must be considered. The primary cost drivers include fine-tuning and certifying the model for this specific purpose and integrating it into hospital systems. The price is also expected to rise in response to the impending future regulation (e.g. EU AI Act).
- **Algorithmic bias:** It can occur when AI algorithms are trained on large datasets that are missing or misrepresenting several groups or populations. This bias can result in misdiagnoses and a lack of generalization[18].
- **Real-Time update:** Medical guidelines and research findings evolve rapidly, and LLMs like ChatGPT may not always reflect the most current data or recommendations, which is critical in fast-paced medical fields like gastroenterology. This could lead to the dissemination of outdated information to healthcare professionals and patients.
- **Distrust:** Both physicians, including gastroenterologists, and patients often express distrust in LLMs. Physicians may be skeptical about the accuracy and reliability of AI-generated recommendations, fearing errors or biases in the data used to train these models. Patients, on the other hand, may worry about the impersonal nature of AI and the potential for misdiagnosis. This distrust highlights the need for transparent, explainable AI systems and thorough validation

processes to ensure that LLMs can be safely and effectively integrated into clinical practice.

Future of LLMs in gastroenterology

One of the key aspects of the future of LLMs in gastroenterology is its ability to process and analyse large amounts of data, helping to identify patterns and create more personalized treatment plans. With advances in AI and deep learning, LLMs will become even more accurate and efficient, allowing for faster diagnosis and treatment of gastroenterological conditions. In the realm of education and training, LLMs will serve as an invaluable resource for information on the latest research articles and procedures, enabling doctors to keep up with the ever-evolving field. On the other hand, the future of LLMs in gastroenterology will also depend on addressing challenges related to data security, privacy, and ethics[19,20]. It is essential to ensure proper protocols and procedures are in place to enable the safe and ethical use of AI in medicine. Critical tasks, especially in data analysis, decision-making, and communication, will still require human intervention, as full automation isn't viable yet. Interface design should promote trust and caution, while also preventing automation from impeding the skill development of medical trainees and practitioners. Overall, LLMs are expected to play an increasingly important role in gastroenterology, contributing to better patient care and supporting doctors in their work[21].

DISCUSSION

AI in medicine opens new horizons and changes the way healthcare is provided, leading to improvements in diagnosis, treatment, and prevention of various diseases. AI helps doctors analyse medical images, such as X-rays, CT scans, and MRI scans, faster and more accurately, thanks to sophisticated algorithms that can detect pathological changes difficult to recognize by the human eye. Zhuang's et al. review demonstrates the power of AI in the field of imaging methods in gastroenterology. It highlights the utilization of AI in endoscopy, sonography, computed tomography, and positron emission tomography. The review also points out the limitations of AI, such as inaccuracies in assessing depth of invasion in endosonographic or erroneous interpretations in cases of inadequate bowel preparation[22]. Recently published article by Wenbo li et al. explores the potential of ChatGPT (specific LLM) in colorectal surgery. It discusses benefits such as enhanced healthcare delivery, personalized information, error reduction, and improved patient outcomes, alongside challenges like patient privacy, liability, and potential effects on doctor-patient relationships. The study suggests that ChatGPT could revolutionize colorectal surgery with its applications[23].

With the ability to process vast amounts of health data, such as electronic health records, genetic information, and clinical trial results, AI enables better predictions of disease progression and selection of the most appropriate treatment methods[24]. Moreover, AI can accelerate research and development of new drugs by helping to identify potential therapeutic targets and optimize chemical structures of drugs, significantly reducing time and costs associated with traditional drug development[25]. Lahat et al. assessed ChatGPT's ability to identify research priorities in gastroenterology, questions were generated on topics like IBD, the microbiome, AI in gastroenterology, and advanced endoscopy. A panel of gastroenterologists rated these questions for importance and relevance, yielding an average score of 3.6 out of 5, with high inter-rater reliability (0.80 to 0.98, $P<0.001$). While the questions scored high for relevance and clarity, they lacked originality, indicating that while LLMs can help pinpoint gastroenterology research priorities, enhancing their novelty remains a challenge[26].

Although AI offers significant potential for improving the quality of care, it is essential to address issues related to ethics, data protection, and collaboration between AI and healthcare professionals to ensure that technology is used responsibly and effectively. World Health Organization published in 2021 first global report on AI with guiding principles for its design and use[27]. Also, the FDA, alongside its various centers, is initiating discussions on leveraging AI and machine learning in drug and medical device development to shape the regulatory landscape. This initiative aims to keep drugs safe and effective while embracing technological advancements in data handling and computation[28].

As mentioned earlier, the use of AI in gastroenterology provides many positives, and we can easily imagine how it can improve life in clinical practice and, especially, in the field of education and patient education[29]. Accelerating diagnosis can help speed up patient treatment, which will also have a favourable economic impact on healthcare. We see significant benefits of AI in accelerating differential diagnosis in patient treatment and, therefore, more efficient decision-making in clinical practice[30]. AI will also enable faster evaluation in endoscopy (detection of polyps, their characteristics, capsule endoscopy reading) and radiology, significantly reducing the risk of human error[31].

On the other hand, we are concerned that the widespread use of AI in practice will eventually lead to an unknowing transfer of responsibility to AI and a certain degree of "dulling" of one's opinion. We will rely on AI in more than 90% of cases, and instead of using our judgment and experience, the first thing we will reach for is AI. This can, of course, lead to incorrect decisions, as we know that AI is not infallible at this time, its knowledge is currently limited, and careful interpretation of AI outputs is required. One of the main ethical issues is ensuring the protection of patients' personal and sensitive health data, which we will make available to AI. It is necessary to carefully balance the benefits of AI with the need to preserve privacy and security of this information. AI developers and healthcare organizations must collaborate to create robust security measures and comply with relevant regulations. The path to integrating such technology seamlessly into medical practice is fraught with ethical and practical challenges. These challenges must be navigated carefully to leverage AI's benefits without compromising patient care or privacy[32]. We believe that informed consent is an integral part of any medical intervention. Patients must be made aware that AI is being used, its purpose, and its limitations. Transparency is key to maintaining trust between the medical community and patients. With technological advances and societal changes, ethics guidelines for AI in medicine need to be regularly reviewed and updated. It is a dynamic field that requires continuous learning and adaptation. To address these ethical challenges effectively, collaboration between ethicists, technologists, clinicians, and policymakers is crucial[33]. The European Union's (EU) proactive stance in developing comprehensive strategies and regulations for artificial intelligence, notably through its proposed AI Act and Ethics Guidelines for Trustworthy AI, sets a vital precedent for responsible AI governance. These initiatives emphasize transparency, safety, and ethical standards, including human oversight, technical robustness, privacy, diversity, societal well-being, and accountability, addressing the profound ethical, legal, and societal implications of AI. In the EU, LLMs used for medical purposes may be classified as Software as a Medical Device, subject to strict requirements under the Medical Device Regulation[34]. This includes rigorous testing and validation to ensure safety and efficacy. There's a clear need for other states and global entities to consider similar frameworks to ensure AI is used in ways that protect citizens' rights and promote trust and safety worldwide. By incorporating these seven key requirements for trustworthy AI, we can foster innovation while ensuring AI benefits society as a whole[35,36].

Another challenge is establishing effective collaboration between AI and healthcare professionals. Doctors, nurses, and other healthcare workers must be able to trust and rely on the AI tools they use and recognize potential errors or shortcomings. This includes providing training and education that enables them to better understand and use AI technologies, as well as ensuring that AI systems are designed to support human interaction, collaboration.

When introducing LLMs into gastroenterological practice, careful consideration must be given to protecting patients' personal data and securing the system. As AI becomes integrated into healthcare, the collection and analysis of patient data are inevitable. This raises the risk of privacy breaches and data misuse. Therefore, strict data security measures should be implemented to ensure patient data privacy, including anonymization and encryption of patient data[37].

Furthermore, the implementation of LLMs involve substantial costs including technology costs, staff training, and system maintenance. It is crucial to evaluate whether these costs outweigh the benefits and efficiency improvements that ChatGPT brings, and to consider whether resources might be better used elsewhere in healthcare provision.

It is also essential to emphasize that LLMs should serve as a complementary tool, not a replacement for human healthcare professionals. The ultimate decision-making power must lie with the humans, who are able to comprehend the nuances and complexities of each individual case.

Ultimately, it is essential to consider the use of LLMs, with the needs and expectations of both doctors and patients in mind. The benefits offered by LLMs should be maximized without compromising the quality of care, security, and patients' trust in the medical system[38].

CONCLUSION

In conclusion, further development in the field of AI can be expected, and we dare say that development will proceed at an exponential speed. In a short time, it will be challenging to evaluate what is created by humans and what is already AI. The future of LLMs in gastroenterology are full of potential and opportunities to improve patient care quality. As an AI tool, LLM will be increasingly integrated into clinical and research environments, where it can support doctors and medical staff in diagnosis, treatment, and education. The development of AI, like ChatGPT and others, is still in full swing, and its capabilities and applications in medicine, including gastroenterology, are expected to continue to grow. However, we must also keep in mind the potential pitfalls of deploying LLMs, ensuring we mitigate any risks associated with its use. Future research and innovation should aim to improve the accuracy, efficiency, and safety of AI in medicine.

Search strategy and selection criteria:

Our search strategy involved using the following databases:

- **Medical:** PubMed, MEDLINE, EMBASE
- **AI/Tech:** IEEE Xplore, ACM Digital Library, arXiv
- **General:** Web of Science, Scopus, Google Scholar

We utilized the following search terms:

- **Core:** "Artificial Intelligence", "Large Language Models", "Gastroenterology"
- **Related:** "Machine Learning", "NLP", "ChatGPT", "Diagnosis", "Treatment"
- **Challenges:** "Bias", "Privacy", "Security", "Cost"

Acknowledgement: Ministry of Health, Czech Republic – conceptual development of research organization, Motol University Hospital, Prague, Czech Republic 00064203.
Author contributions: All authors contributed equally.
Conflict of interest statement: The authors state that there are no conflicts of interest regarding the publication of this article.

REFERENCES

1. Kaul V, Enslin S, Gross SA. History of artificial intelligence in medicine. Gastrointest Endosc 2020;92:807-12. doi: 10.1016/j.gie.2020.06.040
2. Toosi A, Bottino AG, Saboury B, Siegel E, Rahmim A. A Brief History of AI: How to Prevent Another Winter (A Critical Review). PET Clin 2021;16:449-69. doi: 10.1016/j.cpet.2021.07.001
3. Egger J, Gsaxner C, Pepe A, Pomykala KL, Jonske F, Kurz M, Li J, Kleesiek J. Medical deep learning-A systematic meta-review. Comput Methods Programs Biomed 2022;221:106874. doi: 10.1016/j.cmpb.2022.106874

4. Haug CJ, Drazen JM. Artificial Intelligence and Machine Learning in Clinical Medicine, 2023. N Engl J Med 2023;388(13):1201-8. doi: 10.1056/NEJMra2302038
5. Harrer S. Attention is not all you need: the complicated case of ethically using large language models in healthcare and medicine. Ebiomedicine 2023;90:104512. doi: 10.1016/j.ebiom.2023.104512
6. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, Cui C, Corrado G, Thrun S, Dean J. A guide to deep learning in healthcare. Nat Med 2019;25(1):24-9. doi: 10.1038/s41591-018-0316-z
7. Brown T, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, Neelakantan A, Shyam P, Sastry G, Askell A, Agarwal S, Herbert-Voss A, Krueger G, Henighan T, Child R, Ramesh A, Ziegler D, Wu J, Winter C, Amodei D. Language Models are Few-Shot Learners. Arxiv 2020. Available from: https://arxiv.org/abs/2005.14165
8. Naveed H, Khan AU, Qiu S, Saqib M, Anwar S, Usman U, Akhtar N, Barnes N, Mian A. A Comprehensive Overview of Large Language Models. Arxiv 2023. Available from: https://doi.org/10.48550/arXiv.2307.06435
9. Fedus W, B. Z., Shazeer N. Switch Transformers: Scaling to Trillion Parameter Models with Simple and Efficient Sparsity. Arxiv 2021. Available from: https://arxiv.org/abs/2101.03961
10. Cretu C. How Does ChatGPT Actually Work? An ML Engineer Explains. Arxiv 2023. Available from: https://www.scalablepath.com/data-science/chatgpt-architecture-explained
11. Josh A, Adler S, Agarwal S, Ahmad L, Akkaya I, Aleman FL. GPT-4 Technical Report. Arxiv 2023. Available from: https://arxiv.org/abs/2303.08774
12. Bubeck S, Chandrasekaran V, Eldan R, Gehrke J, Horvitz E, Kamar E, Lee P, Lee YT, Li Y, Lundberg S, Nori H, Palangi H, Ribeiro MT, Zhang Y. Sparks of Artificial General Intelligence: Early experiments with GPT-4. Arxiv 2023. Available from: https://doi.org/10.48550/arXiv.2303.12712
13. Alberts IL, Mercolli L, Pyka T, Prenosil G, Shi K, Rominger A, Afshar-Oromieh A. Large language models (LLM) and ChatGPT: what will the impact on nuclear medicine be? Eur J Nucl Med Mol Imaging 2023;50(6):1549-52. doi: 10.1007/s00259-023-06172-w
14. Nori H, King N, Mayer McKinney S, Carignan D, Horvitz E. Capabilities of GPT-4 on Medical Challenge Problems. Arxiv 2023. Available from: https://doi.org:doi.org/10.48550/arXiv.2303.13375
15. OpenAI. Introducing ChatGPT. 2022 Available from: https://openai.com/index/chatgpt/
16. Sharma P, Parasa S. ChatGPT and large language models in gastroenterology. Nat Rev Gastroenterol Hepatol 2023;20(8):481-2. doi: 10.1038/s41575-023-00799-8
17. Sallam M. ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. Healthcare (Basel) 2023;11(6):887. doi: 10.3390/healthcare11060887
18. Norori N, Hu Q, Aellen FM, Faraci FD, Tzovara A. Addressing bias in big data and AI for health care: A call for open science. Patterns (N Y) 2021;2(10):100347. doi: 10.1016/j.patter.2021.100347
19. Shahab O, El Kurdi B, Shaukat A, Nadkarni G, Soroush A. Large language models: a primer and gastroenterology applications. Therap Adv Gastroenterol 2024;17:17562848241227031. doi: 10.1177/17562848241227031
20. Wang F, Preininger A. AI in Health: State of the Art, Challenges, and Future Directions. Yearb Med Inform 2019;28(1):16-26. doi: 10.1055/s-0039-1677908
21. Le Berre C, Sandborn WJ, Aridhi S, Devignes MD, Fournier L, Smaïl-Tabbone M, Danese S, Peyrin-Biroulet L. Application of Artificial Intelligence to Gastroenterology and Hepatology. Gastroenterology 2020;158(1):76-94.e2. doi: 10.1053/j.gastro.2019.08.058
22. Zhuang H, Zhang J, Liao F. A systematic review on application of deep learning in digestive system image processing. Vis Comput 2023;39(6):2207-22. doi: 10.1007/s00371-021-02322-z
23. Li W, Zhang Y, Chen F. ChatGPT in Colorectal Surgery: A Promising Tool or a Passing Fad? Ann Biomed Eng 2023;51(9):1892-7. doi: 10.1007/s10439-023-03232-y
24. Seong D, Choi YH, Shin SY, Yi BK. Deep learning approach to detection of colonoscopic information from unstructured reports. BMC Med Inform Decis Mak 2023;23(1):28. doi: 10.1186/s12911-023-02121-7
25. Barrett JS, Oskoui SE, Russell S, Borens A. Digital Research Environment(DRE)-enabled Artificial Intelligence (AI) to facilitate early stage drug development. Front Pharmacol 2023;14:1115356. doi: 10.3389/fphar.2023.1115356
26. Lahat A, Shachar E, Avidan B, Shatz Z, Glicksberg BS, Klang E. Evaluating the use of large language model in identifying top research questions in gastroenterology. Sci Rep 2023;13(1):4164. doi: 10.1038/s41598-023-31412-2
27. WHO. Growing use of AI for health presents governments, providers, and communities with opportunities and challenges. 2021 Available from: https://www.who.int/news/item/28-06-2021-who-issues-first-global-report-on-ai-in-health-and-six-guiding-principles-for-its-design-and-use
28. FDA. Using artificial intelligence and machine learning in the development of drugs and biological products. Discussion paper and request for feedback. 2023 Available from: https://www.fda.gov/media/167973/download
29. Lee TC, Staller K, Botoman V, Pathipati MP, Varma S, Kuo B. ChatGPT Answers Common Patient Questions About Colonoscopy. Gastroenterology 2023;165(2):509-511.e7. doi: 10.1053/j.gastro.2023.04.033
30. McDuff D, Schaekermann M. Towards Accurate Differential Diagnosis with Large Language Models. ArXiv 2023. Available from: Towards Accurate Differential Diagnosis with Large Language Models
31. Ali S. Where do we stand in AI for endoscopic image analysis? Deciphering gaps and future directions. NPJ Digit Med 202;5(1):184. doi: 10.1038/s41746-022-00733-3
32. De Angelis L, Baglivo F, Arzilli G, Privitera GP, Ferragina P, Tozzi AE, Rizzo C. ChatGPT and the rise of large language models: the new AI-driven infodemic threat in public health. Front Public Health 2023;11:1166120. doi: 10.3389/fpubh.2023.1166120
33. Murdoch B. Privacy and artificial intelligence: challenges for protecting health information in a new era. BMC Med Ethics 202;22(1):122. doi: 10.1186/s12910-021-00687-3
34. Meskó B, Topol EJ. The imperative for regulatory oversight of large language models (or generative AI) in healthcare. NPJ Digit Med 2023;6(1):120. doi: 10.1038/s41746-023-00873-0
35. Commission E. Ethics Guidelines for Trustworthy AI. 2021 Available from: https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html
36. Commission E. A European approach to artificial intelligence. 2023 Available from: https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence
37. Jill M. Security, Privacy Risks of Artificial Intelligence in Healthcare. 2021 Available from: https://healthitsecurity.com/features/security-privacy-risks-of-artificial-intelligence-in-healthcare
38. Clusmann J, Kolbinger FR, Muti HS, Carrero ZI, Eckardt JN, Laleh NG, Löffler CML, Schwarzkopf SC, Unger M, Veldhuizen GP, Wagner SJ, Kather JN. The future landscape of large language models in medicine. Commun Med (Lond) 2023;3(1):141. doi: 10.1038/s43856-023-00370-1