Manuscript Title: Artificial intelligence in gastroenterology: Enhancing clinical practice, managing challenges and exploring future directions.

Manuscript ID: 110109

Corresponding Author: Prof. Arjuna P. de Silva

Response to Reviewer Comments

Reviewer #1

Comment 1: *More fundamental concepts about AI such as the semi-supervised learning, self-supervised learning, and ANNs connecting bridge between ML and DL should be highlighted clearly to help readers better understand what they mean.*

Response: Thank you for this insightful suggestion. We have now included a section in the manuscript that clearly explains AI concepts such as supervised, semi-supervised, self-supervised learning and ANNs, as well as their roles in machine learning and deep learning:

> "Certain ML algorithms are inherently scalable to DL architectures and serve as effective bridges between traditional ML and DL approaches. Notably, semi-supervised learning and self-supervised learning exemplify such strategies. Semi-supervised learning leverages a small amount of labeled data alongside a large volume of unlabeled data to enhance model performance, making it particularly valuable in domains where annotated data is often limited [34]. In contrast, self-supervised learning generates supervisory signals directly from the input data by formulating pretext tasks, allowing models to learn from vast unlabeled datasets. For example, in medical imaging, models can be trained to reconstruct randomly masked regions of chest X-rays. Through this process, the algorithm learns general anatomical features and pathological patterns by predicting missing portions of the image based on the surrounding visible context [35]."

Comment 2: *The authors did not elaborate on the relevant underlying logic of CNNs, for example, how CNN architecture performs semantic segmentation and instance segmentation.*

Response: We appreciate this valuable feedback. We have expanded the section on convolutional neural networks (CNNs) to explain how CNNs are used in medical imaging, particularly for tasks such as semantic segmentation and instance segmentation in gastroenterological applications:

> "CNNs are extensively used for semantic segmentation, where each pixel in an image is classified into predefined categories, such as normal tissue, malignant lesion, or inflamed area, allowing precise identification of pathological regions. This pixel-level classification enables accurate

> mapping of lesion boundaries aiding targeted interventions [26]. CNNs also perform instance segmentation, which not only labels each pixel but also differentiates between distinct instances of the same class (e.g., multiple polyps within a single colonoscopy image). This capability is critical in clinical scenarios where distinguishing individual lesions impacts treatment decisions. Algorithms such as U-Net and Mask R-CNN have been widely used for these segmentation tasks, facilitating enhanced visualization, quantification, and monitoring of disease progression [27]"

Comment 3: *Why does the article not further review the development of RNNs and Transformers for clinical interaction and decision support based on NLPs and LLMs?*

Response: Thank you for the valuable feedback. We have now added content discussing the evolution and application of recurrent neural networks (RNNs) and Transformer-based models, such as BERT and GPT, in natural language processing tasks relevant to gastroenterology, including clinical text mining and decision support:

> "Similarly, RNNs are widely used in natural language processing (NLP) tasks such as analyzing clinical notes, patient histories, and symptom timelines. However, their limitations in capturing long-range temporal dependencies led to the development of Transformer neural networks (TNNs). Transformers utilize self-attention mechanisms to identify and prioritize the most relevant features within input data. These models are applied across clinical document analysis, speech recognition, and even image-based diagnosis. The attention mechanisms can also highlight critical regions in medical images, aiding clinicians in making more accurate diagnostic decisions [23]. Transformers form the foundation of large language models (LLMs) like GPT and BERT, which excel at understanding and generating human-like medical text [24]. These models support advanced clinical decision-making by enabling rapid retrieval of medical knowledge, summarizing vast medical literature, and providing context-aware recommendations. This also significantly improves clinician-patient communication."

Comment 4: *The authors need to clarify the realistic limitations of AI in clinical practice, especially black-box issues and hallucinations in DL.*

Response: Thank you for the valuable feedback. We have expanded the discussion on the limitations of AI in clinical practice, including model opacity (the 'black box' issue), hallucination risks associated with large language models (LLMs), challenges in generalizability across diverse clinical settings, and ethical considerations:

> "Ethical considerations also represent a major obstacle, particularly regarding algorithmic bias, transparency, and accountability. AI tools trained on non-diverse or unrepresentative datasets may yield biased outputs, which can lead to underdiagnosis or misdiagnosis in specific demographic groups, perpetuating health disparities [106] Compounding this issue is the lack of interpretability in many AI models, especially DL architectures. This "black box" nature renders it difficult for clinicians to comprehend or validate the rationale behind AI-generated recommendations, thereby

eroding trust and impeding clinical adoption. Moreover, when such opaque systems make erroneous decisions, attributing responsibility or determining legal liability becomes problematic [107]. Another emerging concern is the phenomenon of AI hallucinations, particularly observed in large language models (LLMs). These models can generate outputs that are syntactically plausible yet factually incorrect or misleading. In a clinical context, such hallucinations pose a serious risk to patient safety if not rigorously vetted by medical professionals [108]."

Comment 5: *More valuable references and unified standards for application, interpretation, and expert comparison should be added.*

Response: Thank you for the valuable feedback. We have revised the manuscript to incorporate recent and relevant references on standards for the application of AI in gastroenterology, as well as appropriate comparisons with expert clinical diagnoses where applicable.

In addition, we have included best practice guidelines for the interpretation of AI-assisted outputs to enhance clinical relevance and utility.

"117 **Glissen Brown JR**, Waljee AK, Mori Y, Sharma P, Berzin TM. Charting a path forward for clinical research in artificial intelligence and gastroenterology. *Dig Endosc* 2022; **34**: 4-12 [PMID: 33715244 DOI: 10.1111/den.13974]

118 **Lekadir K**, Frangi AF, Porras AR, Glocker B, Cintas C, Langlotz CP, Weicken E, Asselbergs FW, Prior F, Collins GS, Kaissis G, Tsakou G, Buvat I, Kalpathy-Cramer J, Mongan J, Schnabel JA, Kushibar K, Riklund K, Marias K, Amugongo LM, Fromont LA, Maier-Hein L, Cerdá-Alberich L, Martí-Bonmatí L, Cardoso MJ, Bobowicz M, Shabani M, Tsiknakis M, Zuluaga MA, Fritzsche MC, Camacho M, Linguraru MG, Wenzel M, De Bruijne M, Tolsgaard MG, Goisauf M, Cano Abadía M, Papanikolaou N, Lazrak N, Pujol O, Osuala R, Napel S, Colantonio S, Joshi S, Klein S, Aussó S, Rogers WA, Salahuddin Z, Starmans MPA; FUTURE-AI Consortium. FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare. *BMJ* 2025; **388**: e081554 [PMID: 39909534 DOI: 10.1136/bmj-2024-081554]"

Editorial Office Comments

Comment 1: *Language classification is Grade C. Please provide a language certificate.*

Response: We acknowledge the recommendation, and the manuscript language has been substantially revised to enhance clarity and flow. If necessary, we are willing to obtain a professional language certificate.

Comment 2: *Manuscript title capitalization*

Response: The title has been corrected to capitalize only the first word, per journal formatting guidelines.

Comment 3: *Author contributions formatting*

Response: Author contributions have been revised to follow the journal's format

Comment 4: *Keywords formatting*

Response: Keywords have been reformatted with capitalized first letters and separated by semicolons.

Comment 5: *Audio core tip*

Response: We have provided the Audio Core Tip.

Comment 6: *Reference formatting issues*

Response: All in-text citations have been changed from "( )" to "[ ]". References have been updated to include PMID, DOI, and all authors. We have used the Auto-Analyser Tool as instructed.

Comment 7: *Figure and table formatting*

Response: All figures and tables have been updated per the specifications, and original figure files have been prepared in PowerPoint. Figure legends now define abbreviations on first appearance. Standard three-line table formatting has also been implemented.

Comment 8: *Image copyright and citations*

Response: All figures included in the manuscript are original and have been appropriately declared within the manuscript