Our response to the first reviewer:

We would like to thank the reviewer for the time spent reviewing our work and for the valuable and thorough feedback. Below we provide our responses to the reviewer's comments in a point-by-point manner:

i. The authors summarized several points in disease diagnostics where the AI and ML techniques can be applied. In-depth discussions will help improve the quality of the manuscript. The manuscript is more suitable to be published in a specialized journal. **Our response:** we would like to thank the reviewer for this comment. We understand the spirit of the reviewer's comment; however, as an invited opinion review, the article's topic was chosen and approved by the journal's editors. Additionally, as per the journal's guide for authors, opinion reviews have a predefined structure and length that we were obliged to follow. Nevertheless, we revised several parts of our manuscript to provide more in-depth discussions for the readers. Changes can be found throughout the manuscript. Additionally, we have added a table in our manuscript summarizing the included studies' findings in an effort to increase readability.

Our response to the second reviewer:

We would like to thank the reviewer for the time spent reviewing our work and for the valuable and thorough feedback. Below we provide our responses to the reviewer's comments in a point-by-point manner:

i. The authors need to correct few typo and grammatical errors or you can take help of English editor services provided by BPG. **Our response:** we would like to thank the reviewer for the comment and apologize for these errors on our end. As per the reviewer's comments, the manuscript was carefully edited by the senior author, who is a native English speaker for any grammatical, semantical/stylistic, and typographical errors. Changes can be found throughout the manuscript.

ii. The authors must add tables depicting sensitivity and specificity of different AI/ML-based Models in Gastroenterology. **Our response:** we would like to thank the reviewer for this comment and apologize for this omission on our end. As per the reviewer's comment, we have now added a table summarizing the parameters used, the classifiers employed, the sample size , the validation strategy, the predicted outcome, and the performance of each study included in our review.

iii. How the predictability and cost-effectiveness can be improved? The authors shall add one para on this. **Our response:** we would like to thank the reviewer for this comment. Regarding the accuracy of predictability we already mention several methodologies to address common challenges such as class imbalance and overfitting and propose a frame of collaboration among physicians and

programmers to avoid shortcomings of the end product. Nevertheless, as per the reviewer's comment, we have added sections elaborating on how to improve predictability accuracy. **The sentences read:** "Standardization of data collection methods is essential to acquire datasets of clean, high-quality data that are representative of a diverse patient population.", "The key is to involve physicians in all the stages of the process. Physicians should be involved in problem identification to ensure that the developed tool addresses actual needs and in data collection and annotation to ensure that the data are labelled correctly, which is crucial in supervised learning, Additionally, we believe that the following steps should be followed during model development to avoid potential drawbacks: physicians should provide a detailed description of the problem at hand to the developing team, explain what is requested by the model (the endpoint), describe the features thoroughly, cooperate to identify and engineer features of high predictive value, discuss the sample size needed based on the nature of the task (binary classification, multi-class classification, regression),ensure that the patients included in the developing procedure are a representative sample of the targeted group, conduct a rigorous clinical validation before the clinical application of the model, provide input on how the tool can be seamlessly be integrated into the existing clinical workflows without causing disruption, discuss the prospects of real-time learning, where the model continues to learn and improves following clinical application, and provide inishts into ethical considerations including ensuring patient privacy and avoiding potential biases. Even following the model's application in the

clinical setting is essential to establish a framework of continuous monitoring and feedback from healthcare professionals to address model shortocomings and improve the model's efficiency. Finally, when applying AI/ML-based tools in the clinical setting is crucial to collect data on the impact of the use of such tools on the clinical outcome, the wellbeing of the patient, and the patient-phycisian relation. Failing to cooperate efficiently and to communicate what exactly is expected from the modelcould result in an end product that does not meet the expectations of the physicians and jeopardizes the safety of the patientsRegarding the improvement of cost-effectiveness, we have now added a section proposing several ways to improve cost-effectiveness**. The section reads:** "There are several ways to improve the cost-effectiveness of implementing AI/ML-based technologies in healthcare. First, a series of models has been developed and patented with a high performance in performing several tasks, such as CAD systems that classify benign from malignant polyps. Thus, employing existing or pre-trained models rather than developing new tools from scratch could significantly decrease the cost of adopting these technologies in clinical practice. When, however, healthcare institutions decide to invest in developing AI/ML-based tools from scratch, a plethora of approaches could be adopted to improve the cost-effectiveness, including the utilization of collaborative platforms and open-source tools, run models on existing hardware when possible to reduce the need of specialized hardware, use of existing datasets when available to avoid data acquisition costs, and finally focus on developing tools that allow for the automatization

of routine, time-consuming tasks while prioritizing high-incidence conditions with the potential for significant cost savings."

iv. The authors need to mention few flow - charts and algorithms of few AI/ML based Softwares commonly used in Gastroenterology. **Our response:** we would like to thank the reviewer for this comment. As per the reviewer's comment, we have added the specific algorithms used by each study included in the review. Additionally, we have added a section in our manuscript describing the more common AI classifiers used in gastroenterology research. **The section reads:** AI classifiers commonly employed in ML tools in gastroenterology are Support Vector Machines (SVM), Artificial Neural Networks (ANN), and Convolutional Neural Networks (CNNS). SVMs are used in supervised learning. Data points the instances are mapped in a high-dimensional space, where the hyperplane that separates the instances based on their assigned label, retaining the highest performance, is selected [22]. SVMs can be used for linear and non-linear problems using kernel functions [23]. ANNs are inspired by the human brain and consist of an input and output layer with at least one hidden layer in between [18]. As the architectures of ANNs became more sophisticated with the addition of multiple hidden layers and even more layer connections, the concept of Deep Neural Networks (DNN) emerged [24,25]. A particular type of DNN, the Convolutional Neural Network (CNN), has found profound application in gastroenterology since the demonstrated high performance in image interpretation. CNNs are based on

convolution, where the image is processed using a series of filters or kernels to detect patterns within the image, such as edges and textures [26].

v. The authors need to mention - what new this study adds to the existing literature? Thanks. **Our response:** we would like to thank the reviewer for this comment and understand its spirit. Nevertheless, as an invited opinion review, this paper does not provide any original content. We were asked to provide our perspective on the current challenges of the application of AI/ML-based technologies in the field of gastroenterology and specifically bring forward 4-5 points of interest and discuss them. Our aim was to discuss important/"traditional" challenges (accuracy, liability, interpretability) but bring forward topics that are not thoroughly discussed in the literature, such as cost-effectiveness and cybersecurity, and elaborate on these topics as much as possible (always within the guidelines provided for an opinion review). However, we are not the first to discuss these points, and we cannot argue in the manuscript that our research brings any novelty in the field of AI/ML application in gastroenterology.

Our response to the third reviewer:

We would like to thank the reviewer for the time spent reviewing our work and for the valuable and thorough feedback. Below, we provide our responses to the reviewer's comments in a point-by-point manner:

i. The accuracy of the developed AI/ML tool is of great importance in the application of AI in the medical field, and in your review, you also spent a lot of time describing the current accuracy status of AI tools and ways to deal with it, so I think accuracy is very necessary to add keywords. **Our response:** we would like to thank the reviewer for this comment. As per the reviewer's comment we have added "accuracy" in the manuscript's keywords.

ii. Emphasis on the importance of physicians: More emphasis on the role of physicians in ML model development and how they can help solve data and model challenges. Provide clear advice on how to involve physicians in the model development process, including advice on the timing of their involvement, how information is delivered, and ensuring sample representation. This can make the article more practical. **Our response:** we would like to thank the reviewer for this comment and agree on its spirit. Physician-programmer collaboration is something we believe is very important and have been mentioned in several parts of our manuscript. As per the reviewer's suggestion, we have elaborated on the section discussing physicians involvement. The sections now reads. "The key is to involve physicians in all the stages of the process. Physicians should be involved in problem identification to ensure that the developed tool addresses actual needs and in

data collection and annotation to ensure that the data are labelled correctly, which is crucial in supervised learning, Additionally, we believe that the following steps should be followed during model development to avoid potential drawbacks: physicians should provide a detailed description of the problem at hand to the developing team, explain what is requested by the model (the endpoint), describe the features thoroughly, cooperate to identify and engineer features of high predictive value, discuss the sample size needed based on the nature of the task (binary classification, multi-class classification, regression),ensure that the patients included in the developing procedure are a representative sample of the targeted group, conduct a rigorous clinical validation before the clinical application of the model, provide input on how the tool can be seamlessly be integrated into the existing clinical workflows without causing disruption, discuss the prospects of real-time learning, where the model continues to learn and improves following clinical application, and provide inishts into ethical considerations including ensuring patient privacy and avoiding potential biases. Even following the model's application in the clinical setting is essential to establish a framework of continuous monitoring and feedback from healthcare professionals to address model shortocomings and improve the model's efficiency. Finally, when applying AI/ML-based tools in the clinical setting is crucial to collect data on the impact of the use of such tools on the clinical outcome, the wellbeing of the patient, and the patient-phycisian relation. Failing to cooperate efficiently and to communicate what exactly is expected from the modelcould result in an end product that does not

meet the expectations of the physicians and jeopardizes the safety of the patients"

iii. In INTERPRETABILITY, while some of the interpretive methods are mentioned, more detailed information can be provided, including how these methods work and how they can help explain the decisions of the black box model. This helps readers better understand how to address interpretative challenges. **Our response:** we would like to thank the reviewer for this comment and apologize for this omission on our end. We have now revised this part of our manuscript as per the reviewer's suggestion to provide an elaborated discussion on the topic. The section now reads: "Introducing a level of interpretability in black boxes applied in healthcare is crucial to enhance patient and physician trust in these tools and facilitate clinical decision-making. Interpretability can be divided into global and local. Global interpretability offers a level of explainability of the model as a whole [50]. Ways to improve global interpretability include feature importance analysis, surrogate models, and interactive visualization tools developed to allow users to explore how the various features influence the model's predictions [50–53]. For example, a model stratifying the risk of CRC development could provide its prediction along with a notification that the prediction is based mainly on the patient's sex, age, and blood count. On the other hand, local interpretability provides the reasoning behind individual predictions [50]. Methodologies to enhance local interpretability include surrogate models, Shapley values, saliency regions, and visualization techniques [50–53]. Shapley values could be used to

demonstrate how each value of each feature contributes to the model's prediction, while saliency maps provide a post-hoc visualization to comprehend which parts of the input were used by the model to reach its decision and are particularly helpful for CNNs to visualize which parts of the image were used in the model's interpretation. For example, a saliency map would highlight exactly which parts of a biopsy the model focused on to reach its diagnosis. Providing a level of explainability for "black-box" models is key to gaining the trust of both physicians and patients and will play a pivotal role in determining which models will dominate the industry."

Our response to the fourth reviewer:

We would like to thank the reviewer for the time spent reviewing our work and for the valuable and thorough feedback. Below we provide our responses to the reviewer's comments in a point-by-point manner:

i. In this viewpoint review, the authors first introduce the current application status of artificial intelligence in the field of gastrointestinal diseases and healthcare. Then, they highlighted a series of challenges faced by these applications, such as accuracy, cost-effectiveness, network security, interpretability, supervision, and accountability. And explored methods to overcome these challenges. As mentioned in the article, what we should pay attention to is: who should be responsible for following the decisions of artificial intelligence models that cause patient harm? The decision-making of the AI/ML model is based on quantifiable parameters. However, clinical doctors' decisions rely on unquantifiable parameters. Artificial intelligence is unlikely to replace doctors. However, it is unlikely that the skills required by future doctors will be similar to those of today. Therefore, doctors should participate in artificial intelligence to avoid becoming outdated. This is a good article with breadth and depth, worth reading for physicians and healthcare workers. **Our response:** we would like to thank the reviewer for the kind comments. We have added several sections in our manuscript, providing a more in-depth discussion of the challenges mentioned.

Our response to the editor:

We would like to thank the editor for the time spent reviewing our work and for the valuable and thorough feedback. Below we provide our responses to the reviewer's comments in a point-by-point manner:

i. This opinion review could be better organized from my point of view. Here are some comments and suggestions: - "Computer systems, in general, and AI/ML tools, particularly, surpass, by far, physicians in their ability to quantify multiple correlations, even in fields where the physicians hold in-depth expertise. Thus, the application of ML tools is particularly useful when the "truth" lies within the data." Even if this is an opinion review, the authors should support the strongest statements. In this case, at least the authors should avoid to reduce the medical act or patients' management to a pure matter of data analysis and calculation. Therefore, I suggest the authors revise their manuscript overall and pay more attention to this aspect or discuss better some aspects. **Our response:** we would like to thank the reviewer for this comment. We tried to organize the "Current Status" section of our manuscript based on: prevention, diagnosis, treatment, and prognosis, and we apologize this was not apparent. We initially tried to reorganize the whole manuscript based on the editor's suggestion (split based on the data type used). However, this drastically changed the manuscript at a stage of the peer review where all language and editing processes have been made. Additionally, the manuscript was provided to us in a format in which our reference managing tool can no longer be used. Therefore, reorganizing the whole "Current Status" at this stage

will introduce a variety of semantic and citation errors. Nevertheless, we have added subtitles to make reinforce the way we organized that particular section of our manuscript. Regarding the comment of supporting our strongest comments, we apologize for this semantic error on our end. We did not intend to insinuate that clinical decisions are based merely on data points. However, these models are usually developed to perform specific tasks (such as flag high risk, interpret a CT scan), rather than make clinical decisions. Nevertheless, we have now revised these parts based on the perspective suggested by the editor. **The sentences now read: "**ML algorithms, on the other hand, perform a series of precise calculations of all the quantifiable variables to perform a certain task" **and** "Thus, the application of ML tools is particularly useful for performing tasks requiring extensive analytical skills such as unraveling correlations in data-sets, following laboratory results trends for long periods of time, and recognizing patterns in various imaging modalities."

ii. - Table 1: it could be more informative. In the last column, the analytical parameter could be abbreviated, but indicate as a note through a number. **Our response:** we would like to thank the reviewer for this comment. As per the editor's suggestion, to increase readability we have revised the last column and used abbreviations for the analytical parameter.

iii. Study design and aims should be better clarified. The authors may change the layout of the word page, in order to add more columns. It is not clear if these studies included a control group. **Our response:** we would like to thank the reviewer for this comment and apologize for this omission on our end. Our original table was in a landscape orientation. The orientation was changed

during the editing of our manuscript by the journal. Nevertheless, as per the editor's suggestion, we have included information regarding control groups and also included the study design of the included studies.

iv. - in order to make this review more objective, I would suggest the authors to re-organize it and create specific sections specifically related to the type of big data that were investigated (endoscopy? molecular aspects? microbiome? others?) **Our response:** we would like to thank the reviewer for this comment. As mentioned above, we have now added to "Current Status" and "Table 1" subtitles to make clear how these studies were organized.