

111902\_Auto\_Edited - 副本.docx

---

WORD COUNT

4676

TIME SUBMITTED

12-NOV-2025 12:39PM

PAPER ID

118994532

**Name of Journal:** *World Journal of Hepatology*

**Manuscript NO:** 111902

**Manuscript Type:** MINIREVIEWS

**Harnessing artificial intelligence in gastroenterology and hepatology: Current applications and future perspectives**

Boutos P *et al.* AI in Gastroenterology and hepatology

**Abstract**

Artificial intelligence (AI) has made remarkable strides, becoming an essential tool in modern medicine. As AI continues to evolve, it is crucial to redefine its scope, classifications, and subtypes to better align with its clinical applications and potential. With a growing number of sophisticated models, AI is now widely used in gastroenterology and hepatology, offering new ways to enhance patient care. In gastroenterology, AI helps doctors identify lesions during endoscopy, detect gastrointestinal bleeding, and support the diagnosis and treatment of conditions like inflammatory bowel disease and gastrointestinal cancers. In hepatology, it aids in staging liver fibrosis, tracking disease progression, and predicting hepatocellular carcinoma risks. Machine learning further personalizes treatment plans, helping physicians make more informed decisions. However, despite its promise, AI still faces hurdles, including biases in data, ethical considerations, regulatory challenges, and the need for better transparency. Moving forward, refining these models, conducting extensive validation studies, and integrating AI seamlessly into clinical practice will be crucial in fully realizing its benefits for gastroenterology and hepatology.

**Key Words:** Trustworthiness; Interpretability; Ethical challenges; Endoscopy; Medical imaging; Clinical decision support; Patient journey; Hepatology; Gastroenterology; Deep learning; Machine learning; Artificial intelligence

Boutos P, Karakasi KE, Katsanos G, Antoniadis N, Kofinas A, Tsoulfas G. Harnessing artificial intelligence in gastroenterology and hepatology: Current applications and future perspectives. *World J Hepatol* 2025; In press

**Core Tip:** This article presents a structured overview of artificial intelligence (AI) in gastroenterology and hepatology, integrating fundamental concepts with real-world clinical applications. By aligning AI tools with each stage of the patient journey, we highlight how AI can enhance prevention, diagnosis, treatment, and follow-up. We also discuss challenges related to trustworthiness, including interpretability, generalisability, and ethics. Through practical examples and original visual frameworks, this article aims to guide clinicians and researchers in understanding, evaluating, and responsibly implementing AI technologies in digestive and liver healthcare.

## **INTRODUCTION**

Artificial intelligence (AI) is rapidly transforming modern medicine, offering new paradigms in data interpretation, decision support, and personalized care. Gastroenterology and hepatology, fields that inherently rely on multimodal data - from endoscopic imaging and histopathology to complex clinical scores and longitudinal biomarkers - are particularly well positioned to benefit from AI-driven innovations. Over the past decade, a surge in machine learning (ML) and deep learning (DL) applications has demonstrated the potential to improve diagnostic accuracy, optimize therapeutic strategies, and enhance risk prediction across a range of digestive and hepatic disorders.

However, the proliferation of AI tools in academic literature has not yet translated into widespread clinical adoption. Significant barriers remain, including technical limitations, ethical concerns, regulatory uncertainty, and lack of integration with existing workflows. Moreover, the diversity of AI methodologies and the often opaque nature of algorithmic decision-making contribute to confusion among clinicians and healthcare stakeholders.

AI refers to a spectrum of computational techniques that emulate human cognitive processes, such as reasoning, pattern recognition, learning, and decision-making. In clinical contexts, AI systems are deployed to interpret complex biomedical data, detect latent structures, and generate actionable insights. AI is typically categorized into “narrow AI”, designed to execute specific clinical tasks - such as polyp classification in endoscopy or fibrosis staging from imaging - and “general AI”, which aspires to approximate human-level intelligence across domains. To date, virtually all medical implementations pertain to narrow AI[1,2].

Translating clinical questions into AI-solvable problems requires formal representation of input data types [*e.g.*, high-resolution endoscopic video, histopathologic slides, longitudinal electronic health record (EHR) sequences] and outcome variables (*e.g.*, disease state, treatment response, or prognosis). Central to the responsible integration of AI are considerations of model interpretability, robustness, and clinical validity. Regulatory frameworks increasingly emphasize model explainability and performance under domain shift - *i.e.*, when applied to populations or settings different from those on which they were trained[3].

Table 1 summarizes the key features of different AI methodologies, including supervised learning, unsupervised learning, reinforcement learning (RL), and large language models. These approaches vary in terms of data supervision, clinical interpretability, and their maturity for deployment in digestive and liver care.

While different AI methodologies differ in structure and application, their translational evolution often follows a similar trajectory. From initial enthusiasm to clinical disappointment, and eventually to mature integration, these

innovations typically pass through defined phases of adoption. This pattern is captured by the well-known “Hype Cycle” model, which helps frame the realistic expectations of AI in medicine over time (Figure 1). Throughout the manuscript, whenever specific AI applications are discussed, the corresponding stage of the Hype Cycle for each technology will be indicated in parentheses. Readers are therefore encouraged to refer back to Figure 1 in order to contextualize these references and recall the sequential flow of the Hype Cycle stages.

This article aims to provide a pragmatic and clinically oriented overview of AI in gastroenterology and hepatology. We follow the patient’s journey through the healthcare system, highlighting AI applications at each stage - from screening and diagnosis to treatment and monitoring. Finally, we address the major challenges in the field, including bias, generalizability, interpretability, and legal-ethical implications, while also offering a forward-looking perspective on responsible AI integration.

By anchoring AI within real clinical pathways and challenges, we hope to foster a deeper understanding of its practical utility and limitations, while empowering clinicians to engage critically with emerging technologies.

### **THE AIPOWERED PATIENT JOURNEY IN GASTROENTEROLOGY AND HEPATOLOGY**

AI is poised to redefine each phase of the patient journey, offering enhanced accuracy, personalization, and efficiency from early risk prediction to long-term disease management. This section systematically examines how AI can be integrated across this continuum, with a specific focus on validated and emerging applications within gastroenterology and hepatology.

#### ***Prevention and early detection***

The preclinical phase of disease management is where AI arguably exerts some of its most transformative effects. By leveraging high-dimensional data - ranging from genetic predispositions and environmental exposures to lifestyle metrics and longitudinal clinical records - AI models can identify at-risk populations with remarkable granularity. It is worth noting that in these screening models the sensitivity and specificity metrics are of great importance.

In the prevention and early detection setting, ML provides the foundation (plateau of productivity). ML, a core subfield of AI, encompasses statistical learning techniques that infer predictive functions from data. Supervised learning remains the most widely adopted paradigm in clinical research, relying on labeled datasets to train discriminative or generative models[4,5]. For instance, annotated colonoscopy frames with polyp labels enable the development of real-time lesion detection tools (slope of enlightenment), while fibrosis scores based on liver biopsy guide ML models for non-invasive fibrosis prediction (peak of inflated expectations).

Unsupervised learning (trough of disillusionment) is instrumental in uncovering latent structures within unlabeled datasets. In hepatology, clustering algorithms have delineated phenotypic subgroups within NAFLD, suggesting differential trajectories of metabolic dysfunction and fibrosis risk. Semi-supervised learning (innovation trigger) leverages small labeled subsets alongside large unlabeled data volumes, an especially advantageous strategy in resource-limited environments. RL, though still in its early clinical applications, is a paradigm in which an agent learns to make sequential decisions through interactions with an environment, with the objective of maximizing a cumulative reward signal. This approach shows considerable promise for dynamic treatment optimization[6,7].

ML pipelines require meticulous attention to preprocessing, feature engineering, algorithm selection, model tuning, and validation. High-stakes medical applications necessitate performance calibration and uncertainty quantification, increasingly evaluated using reliability diagrams and Bayesian ensembling (trough of disillusionment)[8].

In gastroenterology, colorectal cancer screening has become a flagship area for AI deployment. Computer-aided detection systems during colonoscopy (plateau of productivity) have been rigorously validated in randomized controlled trials, with notable improvements in adenoma detection rates (ADR). For instance, the Food and Drug Administration-cleared GI Genius module by Medtronic, trained on over 13 million annotated frames, has been shown to elevate ADR by identifying diminutive and flat lesions often missed by expert endoscopists[9]. A 2022 meta-analysis encompassing 44 randomized controlled trials demonstrated a relative ADR increase from 36.7% to 44.7%, alongside a significant reduction in adenoma miss rates - from 35.3% to 16.1%[10]. In a landmark RCT from Zhejiang University, an AI platform combining RetinaNet (peak of inflated expectations) and ResNet-50 yielded a real-time polyp detection sensitivity of 81.9%, with substantial benefits in identifying flat-type adenomas[11].

In hepatology, ML frameworks have shown immense promise in stratifying risk for hepatocellular carcinoma (HCC), particularly among patients with cirrhosis or viral hepatitis. Ioannou *et al*[12] employed recurrent neural networks to analyze EHR data from over 48000 patients with HCV-related cirrhosis, outperforming conventional regression-based models in forecasting HCC incidence. Likewise, the “SMART” model - a random survival forest incorporating seven longitudinal variables - (slope of enlightenment) has been validated as a robust prognostic tool in patients who have achieved sustained virologic response post-HCV treatment[13].

### *Diagnosis*

The diagnostic process in gastroenterology and hepatology is increasingly augmented by AI algorithms that in many cases outperform traditional image analysis and interpretation methods. In gastrointestinal endoscopy, convolutional neural networks (CNNs) (plateau of productivity) have achieved near-pathologist-level performance.

Here, DL becomes central. DL, a specialized class of ML, is predicated on artificial neural networks with multiple processing layers, enabling the automatic extraction of hierarchical feature representations. CNNs are the predominant architecture for visual data and are widely used in endoscopic image analysis - for example, distinguishing neoplastic from benign colorectal lesions[9]. Similarly, CNNs applied to radiologic imaging support automated liver lesion classification and fibrosis quantification[14].

Temporal data modalities benefit from sequence modeling approaches such as recurrent neural networks, LSTMs (trough of disillusionment), and, more recently, Transformer models that utilize attention mechanisms. Additionally, autoencoders and GANs facilitate dimensionality reduction, anomaly detection, and generation of synthetic clinical data.

DL models are computationally intensive and often data-hungry. Strategies such as transfer learning, multi-task learning, and data augmentation are employed to circumvent data limitations. Interpretability remains a critical bottleneck, prompting the development of explainability frameworks (*e.g.*, Shapley Additive Explanations, Integrated Gradients).

Urban *et al*[4] demonstrated that a CNN could classify colorectal polyps in real time with 96% accuracy (AUC = 0.991), while Byrne *et al*[15] successfully trained a model to distinguish between diminutive adenomas and hyperplastic polyps with performance comparable to histologic assessment. Meta-analytic data suggest that AI systems now surpass human endoscopists in sensitivity (88% *vs* 80%) without compromising specificity[16].

Beyond colonoscopy, AI has revolutionized capsule endoscopy interpretation (Slope of Enlightenment). Architectures such as Focus U-Net (peak of inflated expectations) have achieved Dice similarity coefficients exceeding 0.90, enabling high-throughput identification of bleeding, ulcers, and inflammatory lesions in Crohn's disease[17]. This not only reduces inter-observer variability but also drastically shortens reading times.

In hepatology, AI facilitates non-invasive diagnostics through enhanced image and signal interpretation. DL models trained on elastography and ultrasound datasets have refined fibrosis staging, offering more precise classification of bridging fibrosis and cirrhosis[18]. Furthermore, natural language processing (NLP) (peak of inflated expectations) has proven effective in structuring diagnostic insights from free-text radiology and pathology reports, fostering multidisciplinary coordination and reducing diagnostic latency[19].

Equally critical in the diagnostic domain is computer vision (CV) (slope of enlightenment), which enables automated interpretation of visual data. CV models classify, segment, and detect structures across diverse imaging modalities. In gastroenterology, CV underpins real-time polyp detection systems, quantifies inflammatory burden, and differentiates subtle dysplastic changes. In hepatology, CV supports lesion classification across ultrasound, CT, and MRI, while in pathology it powers whole-slide analysis, grading inflammation, and detecting dysplastic foci with precision that often surpasses inter-observer variability. Domain adaptation techniques are applied to ensure CV algorithms generalize across scanners, staining protocols, and populations.

### ***Prognosis and risk stratification***

Once a diagnosis is established, AI contributes to longitudinal disease modeling, surpassing static clinical scores in predictive granularity.

At this stage, more advanced paradigms like federated (innovation trigger), transfer (innovation trigger), and multi-modal learning (peak of inflated expectations) are most relevant. Federated learning enables decentralized model training across hospitals while preserving privacy. Transfer learning adapts pretrained models to clinical data. Multi-modal learning integrates diverse data streams, uncovering complex phenotypes in conditions like IBD or HCC.

Ethical AI development mandates rigorous attention to bias mitigation, fairness, and transparency. Algorithms must be audited for differential performance across groups, and pipelines should include bias detection steps.

In the realm of IBD, ML models have been constructed to forecast steroid dependence, hospitalizations, and the necessity for surgical intervention. These models synthesize diverse data streams - including serological markers, imaging data, endoscopic severity, and transcriptomics - enabling predictions that outperform standard tools such as the Crohn's Disease Activity Index or Partial Mayo Score[20]. For instance, a gradient boosting model trained on multi-modal inputs has demonstrated superior predictive capacity for 12-month colectomy risk[21].

In hepatology, AI-derived risk models are actively shaping clinical algorithms for decompensation, variceal hemorrhage, and HCC (slope of enlightenment). Models trained on large national registries, including the Veterans Affairs dataset, have exhibited the capacity to predict acute-on-chronic liver failure several weeks prior to its onset[22]. AI is also being employed to recalibrate organ allocation frameworks: By incorporating frailty indices, radiomics, and dynamic laboratory trends into MELD-based models, AI augments the fairness and utility of transplantation prioritization[23].

#### *Treatment decision-making and personalization*

AI is rapidly evolving from a diagnostic adjunct to a decision-support partner capable of informing therapeutic strategy and personalization.

In gastroenterology, RL models (innovation trigger) are being developed to optimize biologic sequencing in IBD, based on probabilistic modeling of long-term remission and adverse events[24]. Additionally, pharmacogenomic AI tools are under development to individualize treatment in *Helicobacter pylori* eradication and gastrointestinal oncology, integrating SNP data and microbial resistance profiles[25].

In hepatology, predictive modeling is being employed to tailor antiviral regimens in hepatitis B and C by analyzing treatment response kinetics and host genomic variables[26]. AI-based models have shown promise in predicting response to locoregional therapies such as transarterial chemoembolization and radiofrequency ablation (slope of enlightenment), thereby aiding modality selection[27]. AI is also being integrated into transplant informatics platforms, with ML models assisting in donor-recipient matching and post-transplant immunosuppressive regimen calibration[28].

### ***Monitoring and long-term management***

Chronic disease monitoring benefits from AI's capacity for continuous data assimilation and anomaly detection.

In IBD, AI-enabled mobile applications aggregate real-time patient-reported outcomes, biomarker data, and wearable sensor outputs to detect subclinical flares or therapeutic failure[34]. These platforms increasingly leverage federated learning approaches, allowing model improvement without centralizing sensitive patient data[29]. For patients in endoscopic surveillance programs (*e.g.*, Barrett's esophagus or post-polypectomy cohorts) (plateau of productivity), AI systems are utilized to standardize follow-up schedules and automate red flag identification during endoscopic review[30].

In hepatology, real-time monitoring of patients with cirrhosis is being augmented through AI algorithms analyzing actigraphy, alcohol intake patterns, and sleep disruption, all of which correlate with hepatic encephalopathy risk[31]. Predictive models applied to lab data trajectories have also demonstrated early warning capabilities for inpatient decompensation events, facilitating pre-emptive outpatient interventions[32].

### ***Patient engagement and telemedicine***

AI-driven tools are also revolutionizing patient engagement, health literacy, and remote care delivery.

In this sphere of engagement and remote care, NLP and conversational AI play a central role. In gastroenterology, conversational AI chatbots provide symptom triage, nutritional advice, and procedural guidance, notably improving bowel preparation adherence[33]. NLP-based systems are increasingly used to generate simplified summaries of procedural and imaging reports, enhancing patient understanding and shared decision-making[34].

In hepatology, intelligent virtual assistants support (peak of inflated expectations) patients with cirrhosis through medication reminders, lifestyle interventions, and psychosocial support. These systems dynamically adapt to patient inputs and behaviors, increasing adherence and engagement[35]. Telemedicine platforms with embedded AI (slope of enlightenment) triage functions have proven effective in monitoring high-risk patients, enabling clinicians to detect and intervene upon early signs of decompensation[36].

Beyond simple chatbots, NLP techniques are evolving toward more advanced patient-facing applications (innovation trigger). Transformer-based language models (*e.g.*, BioBERT, ClinicalBERT, GatorTron) can generate personalized, easy-to-understand summaries of complex medical reports, enhancing health literacy and shared decision-making. NLP also supports real-time translation of medical instructions, enabling non-native speakers or low-literacy patients to better follow therapeutic regimens. Furthermore, sentiment analysis algorithms are beginning to be integrated into patient communication platforms, helping clinicians detect early signs of anxiety, depression, or disengagement in chronic disease management.

Another promising domain is automated documentation: Conversational AI integrated with telemedicine platforms can transcribe, structure, and code virtual consultations in real time, reducing clinician workload while simultaneously generating patient-friendly outputs. These systems are being trialed in gastroenterology follow-up clinics and liver transplant programs, where efficiency and patient comprehension are equally critical.

A structured overview of representative AI applications across the digestive and hepatic patient journey is presented in Table 2. It highlights how AI tools can contribute to each clinical stage, from early risk prediction to post-treatment surveillance.

Collectively, these applications illustrate a paradigm in which AI serves not as a standalone tool, but as an integrated framework augmenting clinical reasoning across all phases of the patient journey. As the evidence base matures, robust validation, seamless workflow integration, and ethical governance will be paramount to realizing AI's full transformative potential - topics explored in the subsequent section.

### **CHALLENGES, LIMITATIONS, AND FUTURE DIRECTIONS IN AI FOR GASTROENTEROLOGY AND HEPATOLOGY**

Despite the transformative potential of AI in the fields of gastroenterology and hepatology, the pathway to seamless integration into everyday clinical workflows is fraught with multifaceted challenges. These limitations extend across technical, ethical, regulatory, operational, and sociocultural domains, necessitating a comprehensive, multidisciplinary, and iterative approach to resolution. A conceptual framework often used to assess the maturity and translational trajectory of AI tools in medicine is the AI readiness pyramid, which delineates a progression from proof-of-concept development, through trustworthiness and clinical validation, to full-scale integration into routine practice (Figure 2). This section provides a rigorous scholarly analysis of the most pressing barriers confronting AI deployment in digestive and hepatic medicine, alongside emerging paradigms, methodological innovations, and strategic frameworks that aim to facilitate responsible adoption and clinical translation.

#### ***Data quality, bias, and generalizability***

The cornerstone of any AI system is the availability of high-quality, representative, and well-annotated data. In gastroenterology and hepatology, the data ecosystem is highly heterogeneous, encompassing diverse modalities such as digital endoscopy, high-resolution cross-sectional imaging (*e.g.*, MRI, CT), histopathological digitization, structured clinical indices (*e.g.*, MELD, Child-Pugh, Mayo Score), and free-text narratives from EHRs[37]. This heterogeneity poses formidable challenges for harmonization, pre-processing, annotation consistency, and downstream reproducibility. Label noise, sampling bias, class imbalance, and inter-observer variability - particularly in histological assessments like fibrosis staging or dysplasia grading - can severely compromise model fidelity and generalizability[38].

Generalization beyond the training domain remains a critical vulnerability. Models trained on data from single institutions or narrow patient demographics frequently demonstrate degraded performance when externally validated across different healthcare systems, geographic regions, or disease subtypes[1]. For instance, CNNs trained to detect cirrhosis using ultrasound imaging from North American datasets may exhibit biased outputs or significant performance attenuation in African or Southeast Asian populations due to differences in etiological prevalence, imaging acquisition protocols, or hepatic phenotypic expression.

Embedded biases - whether racial, gender-based, socioeconomic, or systemic - can be encoded in training datasets, propagating inequities in algorithmic outputs. Addressing these challenges demands deliberate inclusion of diverse datasets, rigorous subgroup performance auditing, and adoption of fairness-aware algorithms. Additionally, advanced methodological approaches such as federated learning, domain adaptation, and transfer learning offer viable strategies to mitigate overfitting, improve domain generalization, and uphold data privacy and regulatory compliance frameworks such as GDPR and HIPAA[15].

### *Interpretability and clinical trust*

One of the most contentious and persistent obstacles in the adoption of clinical AI is the interpretability of complex models, particularly those built upon DL architectures. These systems, often comprising millions or even billions of parameters, generate highly accurate predictions but do so through internal mechanisms that remain largely inscrutable to human observers. As a result, they are frequently described as “black boxes”, since the decision-making pathway cannot be easily traced or rationalized. This lack of transparency poses significant challenges in clinical practice, where trust, accountability, and justification of decisions are indispensable. For instance, when an AI model recommends a diagnostic pathway or flags a high-risk patient, clinicians must be able to explain the reasoning to patients, colleagues, and regulatory bodies. Without such interpretability, even the most statistically robust models risk rejection or underutilization. Moreover, the opacity of DL systems exacerbates ethical concerns, as errors or biases embedded in training data can propagate unchecked, making it difficult to assign responsibility when adverse outcomes occur. Consequently, improving interpretability is not merely a technical ambition but a clinical and ethical necessity. Ongoing research into explainable AI seeks to address this barrier, through methods such as attention mapping, feature attribution, and post-hoc interpretive models, yet these approaches remain imperfect and often oversimplify the underlying complexity. The tension between predictive power and transparency thus constitutes a central dilemma for the future of clinical AI adoption[39]. In critical clinical contexts - such as liver transplantation eligibility or oncologic staging - clinicians demand not only high performance but also an intelligible rationale that can be reconciled with existing clinical heuristics and medical reasoning.

To address this concern, interpretability-enhancing techniques have been developed. <sup>2</sup> Model-agnostic approaches such as Shapley Additive Explanations and Local Interpretable Model-Agnostic Explanations, as well as model-specific methods like Grad- Gradient-weighted Class Activation Mapping and Integrated Gradients, seek to visualize decision

saliency, highlight influential input features, and simulate counterfactual scenarios[45]. While these tools offer valuable insights, their validity, stability, and reproducibility in high-stakes clinical settings remain areas of active investigation.

Building clinical trust in AI extends beyond technical interpretability. It involves participatory co-design frameworks wherein clinicians contribute to iterative model development and refinement. Embedding interpretability from the inception of model construction - rather than treating it as an ex post facto explanation layer - enhances credibility, usability, and safety[40]. Furthermore, real-world monitoring, feedback loops, and continuous learning pipelines are essential to maintain model relevance and clinician trust over time.

### *Regulatory, ethical, and legal considerations*

The regulatory landscape for AI in medicine is in flux, characterized by a lack of harmonization, evolving definitions, and adaptive requirements. <sup>1</sup> Agencies such as the United States Food and Drug Administration, European Medicines Agency, and Health Canada have initiated frameworks for software as a medical device, encompassing premarket validation, real-world evidence generation, and lifecycle surveillance[41]. However, existing regulatory paradigms are largely designed for static software systems, whereas contemporary AI applications - particularly those involving continual learning or self-updating algorithms - challenge these conventional constructs.

To bridge this regulatory mismatch, forward-looking proposals such as algorithmic pre-certification programs, adaptive approval mechanisms, and algorithmic change control protocols are gaining traction. These frameworks advocate for iterative oversight, risk-tiered validation, and ongoing post-deployment evaluation.

Ethically, the integration of AI introduces profound questions around autonomy, data sovereignty, consent granularity, and risk stratification. In hepatology, for instance, AI models that incorporate genetic, lifestyle, or environmental data to predict HCC risk could inadvertently lead to stigmatization, discrimination, or adverse insurance determinations[42].

Ethical AI deployment necessitates rigorous de-identification practices, dynamic consent frameworks, transparent model documentation, and ethical oversight boards.

From a legal perspective, issues of liability in AI-mediated decision-making remain underdefined. Jurisdictional ambiguities exist regarding culpability in cases of algorithmic error - whether responsibility lies with the developer, clinician, institution, or vendor. These ambiguities pose significant deterrents to adoption unless addressed through legal reform, institutional risk management policies, and professional guidelines on AI responsibility.

### *Failures, limitations, and lessons learned*

The evolution of AI in medicine has been punctuated by notable failures, many of which serve as critical learning opportunities. Early efforts to deploy real-time AI-based polyp detection systems in colonoscopy often failed during clinical translation due to insufficient generalizability, poor handling of motion artifacts, operator-induced variability, and lack of integration into endoscopic workflow[21]. These shortcomings underscored the limitations of training on static, curated datasets without accounting for real-world complexity.

In hepatology, predictive models for HCC recurrence frequently exhibit limited applicability across diverse patient populations, especially when trained on homogeneous cohorts or single-etiology data (*e.g.*, HCV-dominant datasets)[14]. Methodological flaws - including data leakage, label imbalance, overfitting, and inadequate external validation - continue to undermine many published models, diminishing clinical confidence.

Methodological rigor is paramount. Adherence to standardized reporting guidelines such as TRIPOD-AI for model development and DECIDE-AI for early clinical evaluation is essential. Prospective multicenter trials, external validation in real-world settings, and robust post-deployment monitoring are prerequisites for responsible implementation.

Furthermore, aligning AI tools with existing clinical workflows and ensuring seamless integration into EHR and PACS systems are indispensable for sustainable impact.

*Future perspectives: Toward a responsible and augmented intelligence*

The future trajectory of AI in gastroenterology and hepatology should prioritize augmentation rather than automation of clinical judgment. Next-generation systems are envisioned to integrate multi-modal data streams - including genomics, proteomics, digital histopathology, imaging biomarkers, and patient-reported outcomes - into context-aware decision support systems that enable personalized, dynamic, and anticipatory care[43].

Human-in-the-loop, or human-supervised AI, refers to AI architectures in which human clinical judgment remains an active and integral component of the decision-making process. In this paradigm, clinicians do not merely serve as passive end-validators but rather as dynamic supervisors who iteratively provide feedback to guide and refine the system, thereby enhancing accuracy and mitigating systemic errors. Such hybrid frameworks aim to merge the computational scalability, speed, and capacity for large-scale data processing inherent to AI with the ethical reasoning, clinical expertise, and contextual sensitivity unique to human practitioners. This synergistic integration aspires not only to improve safety and accountability but also to ensure that algorithmic outputs are meaningfully aligned with the nuanced realities of medical practice.

Foundation models trained on large-scale, cross-specialty, and multi-institutional datasets promise to revolutionize the AI landscape. Fine-tuned versions of such models could serve as universal backbones, supporting a range of clinical tasks from colorectal cancer screening during colonoscopy to advanced fibrosis detection in NAFLD patients using non-invasive imaging. Coupled with remote monitoring platforms and wearable biosensors, AI systems could enable real-time disease tracking and early intervention.

Medical education must evolve to equip future practitioners with core competencies in AI literacy, algorithmic validation, and digital ethics. A useful conceptual framework for situating these stages of development is the “AI readiness pyramid” (Figure 2), which outlines the progressive trajectory from proof-of-concept validation, through the establishment of trustworthiness and clinical testing, to the eventual routine integration of AI into clinical workflows. Beyond general awareness, curricula should emphasize specific skills such as interpreting model validation metrics (*e.g.*, AUC, precision-recall, calibration curves), understanding principles of data privacy and governance, and developing the ability to critically appraise AI-related studies for biases, generalizability, and clinical relevance. Interdisciplinary training programs - spanning clinical medicine, computer science, biostatistics, health informatics, and law - are therefore essential to cultivate the next generation of clinician-data scientists. Finally, the principle of equity must remain foundational. This requires systemic efforts to identify and mitigate bias, involve underserved communities in AI development, and validate models across diverse demographic and geographic cohorts. Ultimately, the promise of AI in digestive and hepatic medicine will be realized not solely through algorithmic sophistication but through principled stewardship, inclusivity, and a relentless commitment to patient-centered innovation.

## **CONCLUSION**

AI is poised to transform gastroenterology and hepatology by enhancing every phase of the patient journey - from early risk prediction and cancer screening to advanced fibrosis staging, treatment personalization, and long-term monitoring of chronic liver and intestinal diseases. In gastroenterology, validated AI tools such as computer-aided detection systems in colonoscopy have already demonstrated substantial improvements in adenoma detection, while in hepatology, ML models are redefining risk stratification for HCC and guiding non-invasive fibrosis assessment. These advances illustrate

the discipline-specific potential of AI to improve diagnostic precision, optimize therapeutic strategies, and standardize follow-up care.

Despite this momentum, widespread implementation remains contingent upon overcoming key challenges: Ensuring the quality and representativeness of multimodal data, improving interpretability of complex models, and embedding AI systems seamlessly into endoscopic suites, hepatology clinics, and transplant programs. Future progress will require robust validation in multicenter trials, alignment with regulatory frameworks, and the development of interdisciplinary training to equip clinicians with AI literacy and evaluative skills.

Ultimately, the promise of AI in gastroenterology and hepatology is not to supplant clinical expertise but to augment it - providing clinicians with powerful, data-driven tools that enhance decision-making, improve patient outcomes, and drive a new era of precision digestive and hepatic medicine.

# 111902\_Auto\_Edited - 副本.docx

---

## ORIGINALITY REPORT

---

1%

SIMILARITY INDEX

---

### PRIMARY SOURCES

---

1 [www.researchgate.net](http://www.researchgate.net)  
Internet

14 words — < 1%

---

2 [pmc.ncbi.nlm.nih.gov](http://pmc.ncbi.nlm.nih.gov)  
Internet

13 words — < 1%

---

EXCLUDE QUOTES ON

EXCLUDE SOURCES < 12 WORDS

EXCLUDE BIBLIOGRAPHY ON

EXCLUDE MATCHES < 12 WORDS