J Bionic Mem 2025; 5(3): 44-62 DOI: 10.53545/jbm.2025.49

Journal of Bionic Memory Review Article

Artificial intelligence in internal medicine: A Comprehensive review

Satilmis Bilgin[1], Sumeyye Buse Balci[1], Tuba Taslamacioglu Duman[1], Burcin Meryem Atak Tel[1], Fatih Baltaci[2], Gulali Aktas[1*]

[1]*Department of Internal Medicine, Bolu Abant Izzet Baysal University Hospital, Bolu, Türkiye*
[2]*Computer Engineer, Ddosify Inc. 1401 Pennsylvania Ave. Unit 105, Wilmington, Delaware 19806, County of New Castle, US*

ABSTRACT

Artificial intelligence (AI) is reshaping the practice of internal medicine by enhancing diagnostic precision, optimizing clinical workflows, and supporting individualized patient care. As digital health technologies mature, AI is increasingly integrated across multiple medical domains, offering new opportunities and challenges for clinicians. This comprehensive review aims to provide an updated overview of the current and emerging applications of AI in internal medicine, highlighting its contributions across major subspecialties such as cardiology, endocrinology, nephrology, gastroenterology, hematology, and oncology. Recent literature demonstrates that AI algorithms, particularly those based on machine learning and deep learning, have achieved notable success in tasks such as medical imaging interpretation, pattern recognition in laboratory and clinical data, and prediction of disease outcomes. In cardiology, AI enhances ECG and echocardiographic analysis; in endocrinology and nephrology, it aids in early detection of diabetic and renal complications; and in oncology and hematology, it supports diagnostic pathology and prognostication. Despite this progress, translation into daily clinical practice remains limited due to challenges related to data quality, model interpretability, generalizability, data safety concerns and ethical considerations. In conclusion, AI holds significant promise to advance internal medicine by augmenting clinical decision making and promoting precision medicine. Real-world integration will require interdisciplinary collaboration, transparent model validation, and regulatory guidance to ensure reliability, safety, and equity. Continued clinical engagement and responsible implementation are essential for transforming AI's technical potential into perceptible benefits for patient care.

Keywords: Internal medicine, endocrinology, cardiology, nephrology, gastroenterology, hematology, oncology, artificial intelligence, large language models, agentic ai.

✉ *Prof. Dr. Gulali Aktas**
Department of Internal Medicine, Bolu Abant Izzet Baysal University Hospital, Bolu, Türkiye
E-mail: draliaktas@yahoo.com
Received: 2025-11-19
Accepted: 2025-12-01 / Published: 2025-12-20

Introduction

In recent years, the landscape of internal medicine has begun to shift under the accelerating influence of artificial intelligence (AI). The growing availability of electronic health records (EHRs), wearable devices, imaging datasets, and other digital health sources has supported the rapid adoption of AI-driven methods, particularly machine learning (ML) and natural language processing (NLP) for diagnostic, prognostic, and therapeutic applications across medicine [1-3]. Within

internal medicine practice, encompassing chronic disease management, multisystem assessment, and complex decision making, the promise of AI is particularly compelling. It would be helpful by assisting in early risk stratification, enabling more individualized treatment choices, and supporting overburdened clinicians. Therefore, AI tools hold the potential to enhance both efficiency and quality of care. For example, surveys of internists reveal that many perceive AI as improving diagnostic accuracy and treatment decisions, though adoption remains uneven across subspecialties [4]. Yet alongside this promise lie substantial challenges. The interpretability of many models with limited transparency, issues of data privacy, algorithmic bias, and workflow integration have been repeatedly cited as significant barriers in internal medicine settings [3,5]. Moreover, despite numerous proof-of-concept studies identified via medical literature, translation to real world implementation in general internal medicine remains limited [6].

Against this backdrop, this review synthesized the medical evidence on AI in internal medicine. We examined the current state of applications, highlighted lessons from deployment, and outlined emerging opportunities as well as critical impediments to broader adoption. The aim is to provide internists, researchers, and policymakers with a clear and actionable overview of how AI is shaping the field today and how it may be meaningfully integrated into internal medicine practice in the coming years.

AI in Endocrinology

Diabetes remains the most extensively studied and best-validated area of AI applications in endocrinology. Artificial intelligence is now embedded across the entire care continuum from early risk prediction (identifying individuals at risk of developing diabetes) and glycemic forecasting, to closed-loop insulin delivery systems (automated insulin pumps), insulin dosing decision support, integration of continuous glucose monitoring (CGM) and wearable data, and advanced patient self-management tools. These approaches have produced measurable improvements in time in range, hypoglycemia reduction, and patient self-management in many studies [7-9].

AI applied to thyroid ultrasound and cytopathology, especially deep learning classifiers, can improve nodule malignancy risk stratification and help reduce unnecessary biopsies and surgeries. Several studies in the literature have shown rapid growth in ultrasound image-based models and have suggested potential cost-effectiveness in selected settings [10-12]. Though prospective data were presented, multicenter validation is still limited in this area. Beyond diabetes mellitus and thyroid disorders, AI research in endocrinology is still in its early stages, with fewer studies and validated applications. AI is increasingly used to build predictive models for metabolic disease complications, including cardiovascular and renal outcomes. Another promising area is the use of algorithms to support test ordering and result interpretation, including optimization of endocrine test panels and improvement of laboratory workflows. Early efforts have applied machine learning to pituitary and adrenal disease diagnostics as well as to rare endocrine tumor phenotyping, but most remain proof-of-concept or retrospective studies. Overall, the literature shows promise but far fewer high-quality, prospective clinical evaluations than in diabetes [13-15].

Common methods include traditional machine learning approaches (random forests, gradient

boosting), deep learning architectures (convolutional neural networks, vision transformers, and hybrid architectures for imaging), and interpretable models designed for clinical deployment. Data sources are electronic health records, laboratory databases, imaging (ultrasound), continuous glucose monitoring and wearables, and, increasingly, multimodal datasets combining genetics and clinical data [13,16].

For diabetes mellitus (closed loop systems, predictive glycemia models), there is good evidence of clinical benefit in trial settings. For most other endocrine applications, evidence is often limited to retrospective cohorts, single-center studies, or algorithmic performance metrics (AUC, sensitivity, and specificity) rather than robust clinical outcome trials. Researchers repeatedly call for prospective, multicenter, and implementation studies to show real-world effectiveness and cost-effectiveness [7,13,17].

Key challenges highlighted in the literature include generalization, explanation, data quality, validation, and clinical integration of those studies on AI use in medicine. Models trained on narrow populations often perform worse when moved to different hospitals, devices, or demographic groups. Moreover, clinicians require interpretable outputs to build trust and support decision-making; however, the uninterpretability of black-box models remains a major barrier to their adoption. In addition, missing data, variable coding, and pooled records limit model performance in the reports in the literature. Regulatory approvals exist for some diabetes mellitus devices, but most endocrine AI tools lack prospective regulatory-grade validation. Consequently, clinical acceptance is limited. Integration of AI into electronic health records and clinical workflows, and the human factors around alerts and recommendations, are frequent barriers [8,10,13].

Practical recommendations for clinicians and researchers may include the definition of validated problems. The adoption of AI should focus on areas where prospective evidence exists, such as regulatory-approved closed-loop insulin delivery systems, and where performance has been validated in large populations. Before local deployment, researchers should insist on model performance reported on external cohorts and subgroup analyses for equity. The researchers and clinicians should prefer interpretable models for decision support where possible. Implementation studies should be planned. They could be built in prospective monitoring of clinical outcomes, workflow impact, patient-reported outcomes, and cost implications.

AI in Nephrology

Artificial intelligence, including supervised, self-supervised, and federated learning approaches, is being actively developed across the nephrology continuum, from early detection and risk prediction (especially for acute kidney injury (AKI) and chronic kidney disease (CKD) progression), through dialysis optimization and remote monitoring, to image-based renal pathology and transplant biopsy interpretation. Recent data in the literature describe rapid growth in both methods and clinical use cases, with the strongest clinical evidence in AKI prediction, dialysis session optimization, and some renal pathology image analysis tasks [18,19]. Predicting AKI earlier than conventional creatinine-based rules is the most mature and visible application. Multiple ML models, including deep learning, trained on large electronic health record datasets, can predict imminent AKI (hours to days ahead), provide dynamic risk scores, and forecast in-

hospital mortality for AKI patients. These models improve identification of at risk patients, enabling earlier monitoring and intervention in retrospective and some prospective evaluations, but prospective clinical outcome evidence of reduced hard endpoints (mortality, dialysis need) remains mixed and limited [20,21]. AI has been used to identify undiagnosed CKD from routine laboratory values and electronic health record signals, to predict which patients will progress to advanced CKD or kidney failure, and to stratify risk for cardiovascular complications. Recent studies highlighted promising discrimination metrics across many models but noted heterogeneity in predictors, cohorts, and outcome definitions, as well as a lack of generalizable external validation. Explainable ML approaches are being explored to increase clinician trust [22,23]. In dialysis, artificial intelligence is being applied to help anticipate intradialytic hypotension, guide ultrafiltration and treatment planning, identify access complications early, and analyze large session datasets to improve care quality. Several studies demonstrated accurate prediction of session level events and point toward precision dialysis systems that adapt parameters in near real time, but most work is algorithmic or retrospective; randomized implementation trials are scarce [19,24]. Moreover, the integration of digital pathology and deep learning has rapidly advanced the detection and quantification of renal lesions, including glomerulus identification, lesion segmentation, fibrosis scoring, and automated lesion classification. Published models can match or exceed pathologists' performance on specific tasks in independent test sets and accelerate workflow. Renal pathology is increasingly recognized as one of the most promising fields for applying image-based AI in clinical practice [25]. Nevertheless, wider validation across centers and consistent standards are needed before such tools can be used routinely. Artificial intelligence has also been applied to predict graft survival and acute rejection, and to interpret transplant biopsies. Recent deep learning studies using whole slide images report strong performance for detecting and subtyping rejection and for identifying features predictive of graft loss, sometimes outperforming local pathologists on technical metrics [25,26]. These developments are promising for earlier and more standardized interpretation of transplant pathology and for graft outcome prognostication, yet prospective validation studies and regulatory approval are still required. Most nephrology AI studies draw on data from electronic health records (including laboratory values, vital signs, and medications), dialysis machine and session logs, imaging and whole-slide pathology, and, less frequently, biomarkers and genomics. Methods range from gradient boosted trees and random forests to recurrent and transformer style deep networks for time series and image convolutional nets for pathology [20,25]. Explanation ability techniques (Shapley Additive Explanations [SHAP] and attention visualization) are increasingly reported to improve interpretability.

AI in Gastroenterology

AI models are being used in endoscopy and lesion detection. Indeed, one of the most advanced applications of AI in gastroenterology is in gastrointestinal (GI) endoscopy. Computer-aided detection (CADe) of polyps in colonoscopy, computer-aided diagnosis and classification (CADx) of lesions, including colorectal polyps, Barrett's esophagus, early gastric cancer, and analysis of capsule endoscopy videos are methods of AI

use in gastroenterology practice [27]. For example, a study noted that AI tools are available for the entire gastrointestinal tract and hepatobiliary system [28]. A meta-analysis found that for the diagnosis of early gastric cancer (EGC) using AI-based endoscopic systems, the pooled area under the curve (AUC) was approximately 0.96 (95% CI, 0.94–0.97), with a sensitivity of about 86% (95% CI, 77–92%) and specificity of about 93% (95% CI, 89–96%) [29]. In a review of registered clinical trials, about half of the AI studies in gastroenterology were related to endoscopy, most often focusing on detecting or classifying colorectal neoplasia [30]. Capsule endoscopy represents an ideal imaging domain for AI applications due to the substantial video data burden, a point emphasized in a recent review of AI in endoscopy [31].

Beyond lesion detection, AI and ML techniques have also been applied to predict outcomes in hepatobiliary diseases, such as survival after hepatocellular carcinoma resection [32]. It is also helpful in predicting treatment response in inflammatory bowel disease, for example, by identifying which patients are most likely to benefit from biologic therapy. [32]. Moreover, it is beneficial for evaluating bowel preparation quality and automatically assessing endoscopic quality metrics [28].

Recently, AI has been applied to hepatobiliary and pancreatic diseases, including imaging-based approaches for liver fibrosis, cirrhosis, hepatocellular carcinoma detection, and pancreatic cyst classification. [33]. Researchers are increasingly testing AI tools to assist with endoscopy workflows, including mucosal feedback, automated quality control, and real-time procedural guidance [28]. These systems may also support remote monitoring in high-risk gastroenterology and hepatology populations, such as patients with cirrhosis who require regular follow up [32].

In colonoscopy, the use of AI for polyp detection has been shown in meta-analyses of randomized controlled trials to reduce adenoma missing rate and increase adenoma detection rate. One meta-analysis concluded that AI assistance significantly reduces miss rates of adenomas and polyps [34]. For early gastric cancer, the systematic review reported very strong diagnostic metrics (AUC about 0.96) for AI in early gastric cancer diagnosis via endoscopy [29]. Reports in literature conclude that AI currently outperforms or matches expert human performance in specific image detection and classification tasks in the GI system [29]. A recent web-based study in the Asia-Pacific region found high acceptance of AI in gastrointestinal endoscopy, with 83% of respondents accepting computer-aided detection and 78.8% accepting computer-aided diagnosis and classification among gastroenterologists and gastrointestinal surgeons. [35].

Despite encouraging advances, many reviews highlight significant limitations. Key challenges include study bias, the black-box nature of models, data heterogeneity, and ethical concerns. Models developed in a single center or within narrow populations often perform less reliably when applied elsewhere, for example, across different devices, patient groups, or imaging modalities [36]. Additionally, clinicians may hesitate to trust AI recommendations without a detailed explanation [33]. Differences in imaging protocols, endoscopic equipment, video quality, and documentation practices often prevent effective data pooling and model development. Ethical questions also arise, particularly around who is responsible when AI-assisted systems contribute to diagnostic

errors in endoscopy. In addition, data privacy, intellectual property, and cybersecurity remain prominent concerns [32,33]. Detection rates rise in trials; however, real-world outcomes, such as reductions in interval cancers and cost savings, are less well established. Also, integration with endoscopy suites and clinician workflows is nontrivial.

In evaluating AI tools for endoscopy, clinicians, researchers, and administrators involved in gastrointestinal practice should give priority to externally validated, multicenter studies with human comparators and prospective trial data. Moreover, they should ensure that the device and AI tool are validated for the specific endoscopy system, patient population, and clinical workflow at the institution. Researchers should prefer tools that provide interpretable outputs or allow human in the loop decision making rather than fully autonomous action. Physicians should remain vigilant for unintended consequences, such as higher false-positive rates, longer procedure times, and alarm fatigue during endoscopy. Outside of endoscopy, including in IBD, hepatology, and pancreatic disease, AI should be considered an adjunctive, exploratory technology at present, since the deployment should be accompanied by monitoring of clinical outcomes and cost outcomes.

AI in Hematology

AI applications in hematology span automated peripheral blood and bone marrow morphology, flow cytometry analysis, digital pathology (whole slide images), genomics and molecular profiling, clinical risk prediction and prognostication, and operational or laboratory automation. Recent studies have documented the rapid expansion of deep learning and other machine learning approaches across these domains, emphasizing that image-rich tasks, such as blood smears, whole-slide imaging, and flow cytometry plots, represent the most mature applications to date [37,38]. Convolutional neural networks and related image recognition pipelines can detect, classify, and count red blood cells, white blood cells, and platelets with accuracy approaching experienced technologists for many tasks, such as WBC differential and schistocyte detection. Large annotated datasets and smartphone microscopy approaches have accelerated model training and potential point of care use [39,40]. Moreover, deep learning is being applied to whole slide bone marrow images for slide level representations, blast quantification, and automated detection of dysplasia or leukemia features. Early studies showed good slide level performance and promise to reduce pathologist workload, but multicenter standardization remains necessary [41]. AI methods have also shown utility in flow cytometry. The technique generates high-dimensional data that are well-suited to machine learning (ML). Recent cross-institutional frameworks and reviews demonstrate that ML can standardize gating, detect abnormal cell populations, and classify conditions such as acute leukemia with very high AUC and accuracy when trained on well-annotated, multisite datasets. Guidance papers now propose risk tiers, validation and revalidation workflows, and governance for clinical flow cytometry ML [42,43].

AI applied to hematoxylin and eosin (H&E) and immunofluorescent staining can assist diagnosis by enabling lymphoma subtyping on biopsies, quantitative assessment of cellularity and fibrosis, and biomarker inference. These image based models can sometimes perform as well as, or even better than, human experts in specific diagnostic tasks. These image based models may match or exceed human performance on narrow tasks and are attractive

for workload triage and objective scoring; however, performance is sensitive to stain and scanner variation and requires robust external validation [44].

ML and deep learning are also used to classify genomic signatures such as subtypes of AML or ALL, to predict prognosis from mutation panels and gene expression, and to aid variant interpretation and therapy matching in hematology practice. Multimodal models combining molecular and clinical data show promise for personalized prognostication and treatment selection, but many are still retrospective and need clinical trial testing [37]. AI is also beneficial for prognosis prediction, risk stratification, and clinical decision making as a supportive tool. Models have been developed to predict outcomes such as relapse, response to therapy, post-transplant graft survival, and the risk of complications, including infection and thrombosis. Several studies reported good discrimination, but the literature often lacks prospective trials showing that model-guided decisions improve patient level outcomes. Implementation studies remain limited [37,45]. In addition, AI is used to triage or prioritize slides for review, reduce false positives, automate cell counts (e.g., hemogram analyzers with onboard image analysis), and enable near-patient testing using finger-prick analyzers and smartphone-based microscopy [39,46]. These tools can increase throughput and reduce human workload when validated appropriately.

Prospective multicenter studies that test whether AI-guided workflows improve meaningful patient outcomes (survival, relapse rates, transfusion needs) are still needed. Standardized benchmarks and external datasets for smear, marrow, and flow cytometry tasks are required to compare models fairly. Multimodal models that integrate imaging, flow cytometry, genomic, and clinical time-series data could enhance precision prognostication and therapy selection.

AI in Cardiology

Artificial intelligence applications in cardiology have progressed considerably, especially in domains supported by large, well-annotated datasets such as electrocardiography (ECG), cardiovascular imaging, device telemetry, and electronic health records (EHRs). Across those domains, deep learning, especially convolutional and recurrent architectures, gradient boosted trees, and multimodal models, are the common approaches. Recent literature highlights the rapid development of technically strong models, with a smaller but growing body of prospective validation studies, clinical implementation research, and regulatory approvals [47-49]. One of the most significant areas is the detection and classification of cardiac arrhythmias. Deep learning models applied to both single-lead and 12-lead ECG recordings have shown high diagnostic accuracy, often matching or even exceeding cardiologists on benchmark datasets. [47,50]. These models power consumer and clinical devices (wearables, ambulatory monitors) and large-scale screening tools. AI can also infer structural or future risk signals, such as detection of reduced ejection fraction, prediction of future atrial fibrillation, or 10 year cardiovascular risk from ECG traces that appear normal to clinicians [51]. Such predictive ECG AI tools are now being trialed at scale in health systems. AI plays a growing role in cardiac imaging. In echocardiography, automated view detection, chamber quantification, and even near complete interpretation pipelines are showing strong cross-site performance in large multicenter studies. AI can accelerate image

interpretation, reduce interobserver variability, and enable screening in point of care contexts [48,52]. Prospective evaluations of AI integration into clinical workflows are increasingly. In coronary angiography via computerized tomography, AI assists coronary plaque quantification and supports non-invasive ischemia assessments, such as CT based fractional flow reserve models, to triage patients more effectively [53]. Moreover, AI reduces segmentation and/or analysis time, standardizes tissue characterization, and is being explored for prognostication from imaging phenotypes, addressing cardiac magnetic resonance imaging (MRI) workforce bottlenecks in cardiac MRI studies. High impact work shows strong diagnostic and/or predictive performance in cardiac MRI tasks [54,55]. AI and ML models are widely applied to predict readmission, mortality, pump failure, and to personalize therapy (device and titration) in patients with heart failure. Systematic reviews show many promising models for 30-day readmission and mortality prediction, but variable methodological quality, frequent lack of external validation, and limited proof that model-guided care changes hard outcomes [56,57]. In interventional cardiology, AI is used for procedural planning (CT-derived planning for percutaneous coronary intervention, coronary arterial bypass grafting), intra-procedural image guidance, and to analyze catheterization laboratory data for quality improvement [53]. Despite intense research activity, clinical adoption requires strong integration into catheter laboratory workflows and regulatory clarity. A growing number of cardiovascular AI tools have obtained Food and Drug Administration (FDA) or other regulatory clearance for specific clinical uses (ECG-based risk tools, automated image quantification, device diagnostics). Analyses of approved devices show diversity in applications but also wide variation in the level of publicly available validation evidence; therefore, clinicians should appraise each tool's external validation and intended use [49,58]. Larger and multicenter randomized or implementation studies that test whether AI-guided actions change clinical outcomes, including mortality, readmissions, and major adverse cardiac events, are required. Standards for transparent reporting, external validation, and real-world performance monitoring are also necessary. There is a growing demand for explainable AI approaches aligned with clinical workflows to support adoption and safe implementation.

AI in Oncology

Deep learning models have substantially improved automated detection, segmentation, and characterization of lesions on CT and/or MRI and mammography; several tools are now approved for clinical use and have shown value in triage and workflow prioritization [59,60]. These image tasks are among the most mature AI applications in cancer care. AI is also applied to whole slide images to identify cancer, grading tumors, quantifying immune and/or inflammatory infiltrates, and even suggesting molecular correlates from hematoxylin-eosin stained slides alone in some settings [61]. Pathology is a rapidly advancing domain where AI is improving diagnostic consistency and unlocking new biomarkers. Machine learning methods that integrate genomics, radiomics, and clinical data are being used to derive predictive biomarkers, such as identifying which patients will respond to immunotherapy or targeted therapies, and to stratify patients for precision treatment [62]. Results are promising, but most studies are retrospective and still need prospective, practice-changing validation. AI is also being

used across the drug development pipeline; target identification, virtual screening, lead optimization, and to optimize trial design and patient selection, shortening timelines and prioritizing candidates for testing [63]. Several industry–AI collaborations illustrate this trend. Meta-analyses and systematic reviews show very high AUCs for specific detection and classification problems, including some cancer image classification tasks [64,65]. AI often matches or exceeds human readers on benchmark datasets for narrowly defined problems. It reduces reading time in radiology and pathology, flags high risk laboratory test results for rapid review, and reduces interobserver variability on scoring tasks [60]. Multimodal AI systems can detect subtle imaging or histologic patterns that correlate with genomic alterations and clinical outcomes, thereby opening new avenues for developing low cost surrogate biomarkers in cancer detection [61].

In the future, prospective, multicenter, randomized clinical trials are needed to determine whether AI-guided decisions improve patient outcomes rather than merely enhance diagnostic accuracy. These trials should be integrated with clinical workflows and governance; clear user interfaces, clinical pathways for action, and monitoring systems for performance drift and safety.

Large Language Models in Internal Medicine

Large Language Models (LLMs) trained on large datasets interpret information from EHR records, clinical notes, and guidelines, and deliver strong results on tasks such as clinical summarization, diagnostic reasoning, and treatment guideline extraction. The most well-known of these are ChatGPT [66] developed by OpenAI, Copilot [67] by Microsoft, Gemini [68] by Google, and LlaMA [69] by Meta. The field of internal medicine, where text-dense data such as clinical notes, discharge summaries, guidelines, drug labels, research articles, and patient messages are abundant, naturally represents the ideal environment for LLMs to operate most effectively.

LLMs are trained mainly on uncontrolled data available online, and the responses provided depend on this data. Therefore, the biggest challenges to the safe use of LLMs in medicine are data bias [70], lack of transparency, data privacy risks [71], and the production of erroneous responses, known as hallucinations. These problems can lead to serious diagnostic errors in medical practice, and therefore, models should only be used after undergoing rigorous clinical validation.

Systematic reviews published in recent years have shown that LLMs are promising in many areas such as clinical documentation, rapid access to information, decision support and patient communication, but can be inaccurate and potentially harmful when used without supervision in real-life conditions [72].
Fine-tuning with domain-specific clinical data plays a key role in increasing model reliability. Instead of merely improving general language fluency, this process helps the model produce accurate and consistent outputs that align with the clinical context in which it is used [73]. In addition, the Retrieval-Augmented Generation (RAG) framework contributes to transparency and traceability by allowing the model to formulate responses based only on recent and validated information sources [74].

Additionally, techniques such as federated learning [75] and differential privacy enable patient data to be processed locally without leaving the institutional environment; however, these methods still cannot provide absolute privacy. For this reason, ensuring the safe and

reliable use of LLMs in clinical practice requires more than technical safeguards alone. It also depends on the establishment of continuous performance monitoring, regulatory frameworks, and robust ethical oversight mechanisms.

Endocrinology

In a study of patients with type 2 diabetes mellitus, 31% of endocrinologists chose metformin for treatment, while 12% preferred the GPT 4 model. However, GPT 4 was more conservative than endocrinologists, preferring metformin less in patients with kidney dysfunction or a history of gastrointestinal distress [76].

Mondal and colleagues compared physicians and the GPT-4 model in type 2 diabetes treatment planning and found that physicians ordered fewer missing medications, while GPT-4 prescribed fewer unnecessary medications. However, they identified safety issues in 16% of treatment plans created with GPT-4 [77].

Another study examined the consistency of GPT-4 in thyroid nodule biopsy decisions. The GPT-4 response was evaluated using different types of prompts, including those aimed at reducing unnecessary biopsies, those based on ATA guidelines, and those based on TI-RADS guidelines. They demonstrated that GPT-4 biopsy recommendation rates varied significantly even for the same clinical scenario, depending on the prompt format used [78]. ThyGPT is a multimodal artificial intelligence model based on LLaMA-3, developed to classify thyroid nodules by evaluating ultrasound images, USG reports, and thyroid guidelines together. The model functions as a clinical "copilot" that can interact with radiologists in natural language to explain diagnostic rationale, thus providing both decision support and interpretability. The study demonstrated that ThyGPT significantly increases radiologists' diagnostic accuracy, reduces unnecessary biopsies, and does so without increasing the risk of missed cancer [79].

LLMs demonstrate strong performance in providing patient information and clear answers to frequently asked questions; in areas such as diabetes and thyroid, the quality of communication and accuracy of explanations often approaches, and in some cases even surpasses, physician levels. However, the picture is more complex for clinical decision support and treatment planning, and they should be viewed as "helpful tools" for treatment planning and drug regulation, which still face experimental and safety concerns.

Nephrology

Nephrology, where patient education, diet, and fluid management, and informed medication use are crucial in conditions such as chronic kidney disease, has become a field where LLMs can be effectively used. LLM-based chatbots can help patients better understand complex treatment regimens, such as dialysis schedules, medication adjustments, and dietary restrictions. Automated documentation and report generation through EHR integration may also reduce physician workload. Moreover, LLM-powered monitoring systems can support early detection and timely alerting for high-risk clinical scenarios such as acute kidney injury (AKI) [80].

A systematic review also evaluated the application of LLMs in four key areas in nephrology; patient education, workflow optimization, dietary guidance, and laboratory data analysis. GPT-4 and other models have demonstrated high accuracy in tasks such as patient information and data interpretation [81].

Gastroenterology

Language models are quite helpful in gastroenterology, particularly for patient education in patients with IBD and chronic liver disease, colonoscopy preparation instructions, diagnostic algorithms, guideline summaries, literature reviews, and scientific manuscript preparation [82].

LLMs have been able to provide mostly guideline-compliant answers to questions about many diseases, including IBD and IBS [83], H. pylori [84], gastroesophageal reflux disease (GERD) [85], colonoscopy and colorectal cancer [86], pancreatic cancer [87], and liver diseases [88].

Oncology and Hematology

Oncology and hematology are among the fields with the heaviest information load due to complex guidelines, rapidly changing treatments, and numerous clinical trials. Studies are being conducted on large language models for clinical question-answering and second-opinion support. LLMs can provide literature-based answers to topics such as staging, treatment plans, and follow-up protocols [89].
Clinical decision-making in oncology requires integrating multiple data types (e.g., text and images). One study developed an autonomous AI agent that integrated GPT-4 with imaging, pathology, and genomic analysis tools to test the accuracy of personalized clinical decisions in oncology. The model achieved 87.2% correct treatment and diagnostic decisions on 20 clinical cases evaluated by experts [90].

Carl et al. evaluated the use of large language models in oncology for medical information, diagnosis, and recommendations. Although LLMs show high potential, they require further validation studies before clinical application due to differences in accuracy, hallucinations, and lack of standard pathology [91].

Cardiology

Brown et al. [92] demonstrated that large language models can be used to integrate social risk factors into predictive tools for estimating 30-day readmission after acute myocardial infarction. Similarly, Dewaswala et al. [93] noted that LLMs can interpret cardiac MRI reports to help the identification of hypertrophic cardiomyopathy. In addition, combining ECG or cardiac MRI findings with clinical text information has been shown to improve patient risk stratification [94]. A UK Biobank study further reported that LLM-based prediction models performed on par with established cardiovascular risk calculators in forecasting major adverse cardiac events over a ten-year follow-up period [95].

One study evaluated ChatGPT's ability to automatically generate echocardiography reports and provide clinical recommendations based solely on echocardiography measurements; the model was shown to be able to produce cardiological reports with 85.7% accuracy [96]. This finding suggests that LLMs have the potential to reduce clinician workload and accelerate clinical processes by providing diagnostic and follow-up recommendations.

While unimodal AI models are particularly effective in detecting conditions such as cardiac amyloidosis, ejection fraction abnormalities, and atrial fibrillation from 12-lead ECGs or medical images, their performance is constrained by the reliance on a single data modality. On the contrary, multimodal models can jointly process ECG signals and clinical text, allowing them to learn cross-modal patterns and generate clinically meaningful representations without depending on labeled datasets. When paired with large language models (LLMs) such as ChatGPT or LLaMA, these systems have shown superior performance in tasks including summarizing

electronic health records, generating clinical documentation, and supporting physician–patient communication [97].

Another study emphasized that LLM performance can be inconsistent across different clinical scenarios, particularly in acute care settings where sensitivity is critical. It was also noted that the existing literature is primarily based on retrospective data, and prospective evidence based on patient outcomes is quite limited [98].

LLMs have shown notable benefits in text-based clinical workflows. They can help with drafting discharge summaries and consultation letters, summarizing clinical guidelines, conducting literature searches, communicating with patients, and improving health literacy. In these areas, they can increase both efficiency and accessibility. However, their accuracy decreases in situations that require highly complex medical reasoning, such as the management of narrow-therapeutic-index drugs, decision-making in patients with multiple comorbidities, and the evaluation of rare diseases. A primary concern is their tendency to produce confident but incorrect statements without proper sourcing (hallucination), which limits their reliability.

For this reason, LLMs should not be viewed as an authority replacing internal medicine specialists. When used appropriately, they are best seen as support tools that reduce workload, save time, and assist clinical decision-making. At the same time, the role of traditional text-only LLMs is expanding with the development of agentic AI architectures. These systems can combine information from different sources, such as MRI/CT image analysis, laboratory results, and clinical data. Thus, the system is evolving from a structure that merely "generates answers" to a more active clinical support component capable of utilizing external tools and validation steps, aiming to reduce the risk of error.

However, these capabilities are still experimental. Due to ongoing uncertainties regarding ethical considerations, data privacy, legal responsibility, and clinical accountability, positioning LLM-based systems in a supportive role, under the supervision of the expert clinician, rather than at the center of the decision-making process, is currently considered the safest approach.

Conclusion

In internal medicine, recent advances in artificial intelligence are redefining how clinicians diagnose, monitor, and manage disease. Across subspecialties, from cardiology and endocrinology to oncology and nephrology, AI driven tools are demonstrating the ability to analyze complex datasets, uncover hidden patterns, and support earlier, more accurate clinical decision making. The use of deep learning in imaging, laboratory testing, electronic health records, and predictive models supports a more personalized and data-driven approach to patient care. Though its routine clinical use remains in early stages. Major challenges, such as issues of data quality and interoperability, limited external validation, performance bias, data privacy concerns, and the need for transparent models that clinicians can trust, remain. Ethical and regulatory approaches must also progress to ensure these systems support clinical judgment and serve patients fairly. The progress of AI in internal medicine will rely on close collaboration among clinicians, data scientists, and health policy makers. Prospective, multicenter validation studies, transparent reporting standards, and human in the loop implementation strategies will be critical for realizing the full potential of AI. When developed and deployed responsibly, AI can

play a supportive role by augmenting clinical expertise, improving efficiency, and advancing the quality and precision of patient care in internal medicine.

Funding: *The author(s) received no financial support for the research, authorship, and/or publication of this article.*

Conflict of Interest: *The authors declare that they have no conflict of interest.*

References

[1]Barba T, Robert M, Hot A. [Artificial intelligence in healthcare: A survival guide for internists]. Rev Med Interne 2025;46(10):594-600.

[2]Aravazhi PS, Gunasekaran P, Benjamin NZY, Thai A, Chandrasekar KK, Kolanu ND, et al. The integration of artificial intelligence into clinical medicine: Trends, challenges, and future directions. Dis Mon 2025;71(6):101882.

[3]Gratzer K, Servy H, Chiche L. [Artificial intelligence: Guidelines for internists]. Rev Med Interne 2020;41(3):189-91.

[4]Mansoor MA, Ibrahim AF, Kidd N. The Impact of Artificial Intelligence on Internal Medicine Physicians: A Survey of Procedural and Non-procedural Specialties. Cureus 2024;16(9):e69121.

[5]Eckstein J. [Artificial intelligence in internal medicine : From the theory to practical application in practices and hospitals]. Inn Med (Heidelb) 2023;64(11):1017-22.

[6]Sim JA, Huang X, Horan MR, Stewart CM, Robison LL, Hudson MM, et al. Natural language processing with machine learning methods to analyze unstructured patient-reported outcomes derived from electronic health records: A systematic review. Artif Intell Med 2023;146:102701.

[7]Guan Z, Li H, Liu R, Cai C, Liu Y, Li J, et al. Artificial intelligence in diabetes management: Advancements, opportunities, and challenges. Cell Rep Med 2023;4(10):101213.

[8]Mackenzie SC, Sainsbury CAR, Wake DJ. Diabetes and artificial intelligence beyond the closed loop: a review of the landscape, promise and challenges. Diabetologia 2024;67(2):223-35.

[9]Makroum MA, Adda M, Bouzouane A, Ibrahim H. Machine Learning and Smart Devices for Diabetes Management: Systematic Review. Sensors (Basel) 2022;22(5).

[10]Toro-Tobon D, Loor-Torres R, Duran M, Fan JW, Singh Ospina N, Wu Y, et al. Artificial Intelligence in Thyroidology: A Narrative Review of the Current Applications, Associated Challenges, and Future Directions. Thyroid 2023;33(8):903-17.

[11]Sant VR, Radhachandran A, Ivezic V, Lee DT, Livhits MJ, Wu JX, et al. From Bench-to-Bedside: How Artificial Intelligence is Changing Thyroid Nodule Diagnostics, a Systematic Review. J Clin Endocrinol Metab 2024;109(7):1684-93.

[12] Bodoque-Cubas J, Fernández-Sáez J, Martínez-Hervás S, Pérez-Lacasta MJ, Carles-Lavila M, Pallarés-Gasulla RM, et al. Integrating Artificial Intelligence in Thyroid Nodule Management: Clinical Outcomes and Cost-Effectiveness Analysis. J Clin Endocrinol Metab 2025.

[13] Giorgini F, Di Dalmazi G, Diciotti S. Artificial intelligence in endocrinology: a comprehensive review. J Endocrinol Invest 2024;47(5):1067-82.

[14] Gruson D, Dabla P, Stankovic S, Homsak E, Gouget B, Bernardini S, et al. Artificial intelligence and thyroid disease management: considerations for thyroid function tests. Biochem Med (Zagreb) 2022;32(2):020601.

[15] Oikonomakos IT, Anjana RM, Mohan V, Steenblock C, Bornstein SR. Recent advances in artificial intelligence-assisted endocrinology and diabetes. Exploration of Endocrine and Metabolic Diseases 2024;1(1):16-26.

[16] Iftikhar M, Saqib M, Qayyum SN, Asmat R, Mumtaz H, Rehan M, et al. Artificial intelligence-driven transformations in diabetes care: a comprehensive literature review. Ann Med Surg (Lond) 2024;86(9):5334-42.

[17] Parab R, Feeley JM, Valero M, Chadalawada L, Garcia GP, Kar SS, et al. Artificial Intelligence in Diabetes Care: Applications, Challenges, and Opportunities Ahead. Endocr Pract 2025.

[18] Singh P, Goyal L, Mallick DC, Surani SR, Kaushik N, Chandramohan D, et al. Artificial Intelligence in Nephrology: Clinical Applications and Challenges. Kidney Med 2025;7(1):100927.

[19] Kotanko P, Zhang H, Wang Y. Artificial Intelligence and Machine Learning in Dialysis: Ready for Prime Time? Clin J Am Soc Nephrol 2023;18(6):803-5.

[20] Rehman AU, Neyra JA, Chen J, Ghazi L. Machine learning models for acute kidney injury prediction and management: a scoping review of externally validated studies. Crit Rev Clin Lab Sci 2025;62(6):454-76.

[21] Lin Y, Shi T, Kong G. Acute Kidney Injury Prognosis Prediction Using Machine Learning Methods: A Systematic Review. Kidney Med 2025;7(1):100936.

[22] Sabanayagam C, Banu R, Lim C, Tham YC, Cheng CY, Tan G, et al. Artificial intelligence in chronic kidney disease management: a scoping review. Theranostics 2025;15(10):4566-78.

[23] Gogoi P, Valan JA. Machine learning approaches for predicting and diagnosing chronic kidney disease: current trends, challenges, solutions, and future directions. Int Urol Nephrol 2025;57(4):1245-68.

[24] Nobakht E, Raru W, Dadgar S, El Shamy O. Precision Dialysis: Leveraging Big Data and Artificial Intelligence. Kidney Med 2024;6(9):100868.

[25] Huo Y, Deng R, Liu Q, Fogo AB, Yang H. AI applications in renal pathology. Kidney Int 2021;99(6):1309-20.

[26] Seyahi N, Ozcan SG. Artificial intelligence and kidney transplantation. World J Transplant 2021;11(7):277-89.

[27] Nawab K, Athwani R, Naeem A, Hamayun M, Wazir M. A Review of Applications of Artificial Intelligence in Gastroenterology. Cureus 2021;13(11):e19235.

[28] Kröner PT, Engels MM, Glicksberg BS, Johnson KW, Mzaik O, van Hooft JE, et al. Artificial intelligence in gastroenterology: A state-of-the-art review. World J Gastroenterol 2021;27(40):6794-824.

[29] Jiang K, Jiang X, Pan J, Wen Y, Huang Y, Weng S, et al. Current Evidence and Future Perspective of Accuracy of Artificial Intelligence Application for Early Gastric Cancer Diagnosis With Endoscopy: A Systematic and Meta-Analysis. Front Med (Lausanne) 2021;8:629080.

[30] Koleth G, Emmanue J, Spadaccini M, Mascagni P, Khalaf K, Mori Y, et al. Artificial intelligence in gastroenterology: Where are we heading? Endosc Int Open 2022;10(11):E1474-e80.

[31] Hann A, Meining A. Artificial Intelligence in Endoscopy. Visc Med 2021;37(6):471-5.

[32] Mehta Y, Mehta S, Bhayani V, Parikh S, Mehta R. Gastroenterology in the age of artificial intelligence: Bridging technology and clinical practice. World J Gastroenterol 2025;31(36):110549.

[33] Christou CD, Tsoulfas G. Challenges and opportunities in the application of artificial intelligence in gastroenterology and hepatology. World J Gastroenterol 2021;27(37):6191-223.

[34] Jin XF, Ma HY, Shi JW, Cai JT. Efficacy of artificial intelligence in reducing miss rates of GI adenomas, polyps, and sessile serrated lesions: a meta-analysis of randomized controlled trials. Gastrointest Endosc. 2024;99(5):667-75.e1.

[35] Goh WW, Chia KY, Cheung MF, Kee KM, Lwin MO, Schulz PJ, et al. Risk Perception, Acceptance, and Trust of Using AI in Gastroenterology Practice in the Asia-Pacific Region: Web-Based Survey Study. Jmir ai 2024;3:e50525.

[36] Cao JS, Lu ZY, Chen MY, Zhang B, Juengpanich S, Hu JH, et al. Artificial intelligence in gastroenterology and hepatology: Status and challenges. World J Gastroenterol. 2021;27(16):1664-90.

[37] Wang J. Deep Learning in Hematology: From Molecules to Patients. Clin Hematol Int 2024;6(4):19-42.

[38] Basak K, Ozyoruk KB, Demir D. Whole Slide Images in Artificial Intelligence Applications in Digital Pathology: Challenges and Pitfalls. Turk Patoloji Derg 2023;39(2):101-8.

[39] Huang B, Kang L, Tsang VTC, Lo CTK, Wong TTW. Deep learning-assisted smartphone-based quantitative microscopy for label-free peripheral blood smear analysis. Biomed Opt Express 2024;15(4):2636-51.

[40] Shams UA, Javed I, Fizan M, Shah AR, Mustafa G, Zubair M, et al. Bio-net dataset: AI-based diagnostic solutions using peripheral blood smear images. Blood Cells Mol Dis 2024;105:102823.

[41] Mu Y, Tizhoosh HR, Dehkharghanian T, Campbell CJV. Whole slide image representation in bone marrow cytology. Comput Biol Med. 2023;166:107530.

[42] Wang YF, Chu EP, Lin FC, Chen HY, Chen TC, Hanson J, et al. A machine learning framework for cross-institute standardized analysis of flow cytometry in differentiating acute myeloid leukemia from non-neoplastic conditions. Comput Biol Med. 2025;193:110394.

[43] Spies NC, Rangel A, English P, Morrison M, O'Fallon B, Ng DP. Machine Learning Methods in Clinical Flow Cytometry. Cancers (Basel) 2025;17(3).

[44] 44. Zhang J, Choi H, Kim Y, Park J, Cho S, Kim E, et al. Artificial intelligence-based digital pathology using H&E-stained whole slide images in immuno-oncology: from immune biomarker detection to immunotherapy response prediction. J Immunother Cancer. 2025;13(8).

[45] El Alaoui Y, Elomri A, Qaraqe M, Padmanabhan R, Yasin Taha R, El Omri H, et al. A Review of Artificial Intelligence Applications in Hematology Management: Current Practices and Future Prospects. J Med Internet Res. 2022;24(7):e36490.

[46] Bachar N, Benbassat D, Brailovsky D, Eshel Y, Glück D, Levner D, et al. An artificial intelligence-assisted diagnostic platform for rapid near-patient hematology. Am J Hematol. 2021;96(10):1264-74.

[47] Ansari Y, Mourad O, Qaraqe K, Serpedin E. Deep learning for ECG Arrhythmia detection and classification: an overview of progress for period 2017-2023. Front Physiol. 2023;14:1246746.

[48] Holste G, Oikonomou EK, Tokodi M, Kovács A, Wang Z, Khera R. Complete AI-Enabled Echocardiography Interpretation With Multitask Deep Learning. JAMA. 2025;334(4):306-18.

[49] Sardar P, Dudorova E, Chatterjee S, Parikh SA, Brilakis ES, Singh JP. Analysis of FDA-Approved Artificial Intelligence and Machine Learning-Enabled Cardiovascular Devices. JACC Adv. 2025;4(10 Pt 2):102174.

[50] Di Costanzo A, Spaccarotella CAM, Esposito G, Indolfi C. An Artificial Intelligence Analysis of Electrocardiograms for the Clinical Diagnosis of Cardiovascular Diseases: A Narrative Review. J Clin Med. 2024;13(4).

[51] Bartusik-Aebisher D, Rogóż K, Aebisher D. Artificial Intelligence and ECG: A New Frontier in Cardiac Diagnostics and Prevention. Biomedicines. 2025;13(7).

[52] Hirata Y, Kusunose K. AI in Echocardiography: State-of-the-art Automated Measurement Techniques and Clinical Applications. Jma j. 2025;8(1):141-50.

[53] Rivera Boadla ME, Sharma NR, Varghese J, Lamichhane S, Khan MH, Gulati A, et al. Multimodal Cardiac Imaging Revisited by Artificial Intelligence: An Innovative Way of Assessment or Just an Aid? Cureus. 2024;16(7):e64272.

[54] Jiménez-Jara C, Salas R, Díaz-Navarro R, Chabert S, Andia ME, Vega J, et al. AI Applied to Cardiac Magnetic Resonance for Precision Medicine in Coronary Artery Disease: A Systematic Review. J Cardiovasc Dev Dis. 2025;12(9).

[55] Wang YJ, Yang K, Wen Y, Wang P, Hu Y, Lai Y, et al. Screening and diagnosis of cardiovascular disease using artificial intelligence-enabled cardiac magnetic resonance imaging. Nat Med. 2024;30(5):1471-80.

[56] Hajishah H, Kazemi D, Safaee E, Amini MJ, Peisepar M, Tanhapour MM, et al. Evaluation of machine learning methods for prediction of heart failure mortality and readmission: meta-analysis. BMC Cardiovasc Disord. 2025;25(1):264.

[57] Yu MY, Son YJ. Machine learning-based 30-day readmission prediction models for patients with heart failure: a systematic review. Eur J Cardiovasc Nurs. 2024;23(7):711-9.

[58] Hussain A, Guni A, Gandhewar R, Warner-Levy J, Davidson A, Shah K, et al. Is the FDA regulation of cardiology AI devices supporting cardiovascular innovation: a scoping review. Heart 2025. DOI: 10.1136/heartjnl-2025-326307.

[59] Najjar R. Redefining Radiology: A Review of Artificial Intelligence Integration in Medical Imaging. Diagnostics (Basel). 2023;13(17).

[60] Lotter W, Hassett MJ, Schultz N, Kehl KL, Van Allen EM, Cerami E. Artificial Intelligence in Oncology: Current

Landscape, Challenges, and Future Directions. Cancer Discov. 2024;14(5):711-26.

[61] Wang J, Wang T, Han R, Shi D, Chen B. Artificial intelligence in cancer pathology: Applications, challenges, and future directions. Cytojournal. 2025;22:45.

[62] Prelaj A, Miskovic V, Zanitti M, Trovo F, Genova C, Viscardi G, et al. Artificial intelligence for predictive biomarker discovery in immuno-oncology: a systematic review. Ann Oncol. 2024;35(1):29-65.

[63] Sarvepalli S, Vadarevu S. Role of artificial intelligence in cancer drug discovery and development. Cancer Lett. 2025;627:217821.

[64] Tandon R, Agrawal S, Rathore NPS, Mishra AK, Jain SK. A systematic review on deep learning-based automated cancer diagnosis models. J Cell Mol Med. 2024;28(6):e18144.

[65] Silva H, Santos GNM, Leite AF, Mesquita CRM, Figueiredo PTS, Stefani CM, et al. The use of artificial intelligence tools in cancer detection compared to the traditional diagnostic imaging methods: An overview of the systematic reviews. PLoS One. 2023;18(10):e0292063.

[66] https://openai.com/chatgpt

[67] https://www.microsoft.com/en-us/microsoft-copilot

[68] https://gemini.google.com/

[69] https://llama.meta.com/ (2024

[70] Guo, Yufei, et al. "Bias in large language models: Origin, evaluation, and mitigation." arXiv preprint arXiv:2411.10915 (2024)

[71] Chen, Kang, et al. "A survey on privacy risks and protection in large language models." Journal of King Saud University Computer and Information Sciences. 2025;37(7): 163.

[72] Meng, Xiangbin, et al. "The application of large language models in medicine: A scoping review." Iscience 2024;2(7).

[73] Wu, Xiao-Kun, et al. "Llm fine-tuning: Concepts, opportunities, and challenges." Big Data and Cognitive Computing. 2025;9(4): 87.

[74] Yang, Rui, et al. "Retrieval-augmented generation for generative artificial intelligence in health care." npj Health Systems. 2.1 (2025): 2.

[75] Yao, Yuhang, et al. "Federated large language models: Current progress and future directions." arXiv preprint arXiv: 2024;2409.15723.

[76] Flory, James H., et al. "Large language model GPT-4 compared to endocrinologist responses on initial choice of glucose-lowering medication under conditions of clinical uncertainty." Diabetes Care. 2025; 48 (2): 185-192.

[77] Mondal, Agnibho, et al. "Evaluating the Performance and Safety of Large Language Models in Generating Type 2 Diabetes Mellitus Management Plans: A Comparative Study With Physicians Using Real Patient Records." Cureus. 2025;17.3.

[78] Flory, James H., Aleksandr Petrov, and Michael Tuttle. "MON-391 The Large Language Model GPT-4 Compared to Guidelines on Thyroid Nodule Management Under Conditions of Clinical Uncertainty." Journal of the Endocrine Society 9.Supplement_1. 2025; bvaf149-2166.

[79] Yao, Jincao, et al. "Multimodal GPT model for assisting thyroid nodule diagnosis and management." npj Digital Medicine. 2025;8.1: 245.

[80] Hu, Yongzheng, Jianping Liu, and Wei Jiang. "Large language models in nephrology: applications and challenges in chronic kidney disease management." Renal Failure 2025;47(1): 2555686.

[81] Unger, Zoe, et al. "Clinical applications and limitations of large language models in nephrology: a systematic review." Clinical Kidney Journal. 2025;18(9): sfaf243.

[82] Maida, Marcello, et al. "The application of large language models in gastroenterology: a review of the literature." Cancers. 2024;16(19): 3328.

[83] Kerbage, Anthony, et al. "Accuracy of ChatGPT in common gastrointestinal diseases: impact for patients and providers." Clinical Gastroenterology and Hepatology. 2024;22(6): 1323-1325.

[84] Lai, Yongkang, et al. "Exploring the capacities of ChatGPT: A comprehensive evaluation of its accuracy and repeatability in addressing helicobacter pylori-related queries." Helicobacter. 2024;29(3): e13078.

[85] Henson, Jacqueline B., et al. "Evaluation of the potential utility of an artificial intelligence chatbot in gastroesophageal reflux disease management." Official journal of the American College of Gastroenterology|. ACG. 2022; 10-14309.

[86] Tariq, Raseen, Sheza Malik, and Sahil Khanna. "Evolving landscape of large language models: an evaluation of ChatGPT and Bard in answering patient queries on colonoscopy." Gastroenterology. 2024;166(1): 220-221.

[87] Moazzam, Zorays, et al. "Quality of ChatGPT responses to questions related to pancreatic cancer and its surgical care." Annals of Surgical Oncology. 2023;30.11: 6284-6286.

[88] Yeo, Yee Hui, et al. "Assessing the performance of ChatGPT in answering questions regarding cirrhosis and hepatocellular carcinoma." Clinical and Molecular Hepatology. 2023;29.3: 721.

[89] Carl, Nicolas, et al. "Large language model use in clinical oncology." NPJ Precision Oncology. 2024;8(1): 240.

[90] 90. Ferber, Dyke, et al. "Development and validation of an autonomous artificial intelligence agent for clinical decision-making in oncology." Nature Cancer. 2025; 1-13.

[91] Carl, Nicolas, et al. "Large language model use in clinical oncology." NPJ Precision Oncology. 2024;8(1): 240.

[92] Brown JR, Ricket IM, Reeves RM, et al. Information Extraction From Electronic Health Records to Predict Readmission Following Acute Myocardial Infarction: Does Natural Language Processing Using Clinical Notes Improve Prediction of Readmission? J Am Heart Assoc. 2022;11(7):e024198.

[93] Dewaswala N, Chen D, Bhopalwala H, et al. Natural language processing for identification of hypertrophic cardiomyopathy patients from cardiac magnetic resonance reports. BMC Med Inform Decis Mak. 2022;22(1):272.

[94] Boonstra, Machteld J., et al. "Artificial intelligence: revolutionizing cardiology with large language models." European Heart Journal. 2024;45(5): 332-345.

[95] Han, Changho, et al. "Evaluation of GPT-4 for 10-year cardiovascular risk prediction: insights from the UK Biobank and KoGES data." Iscience. 2024;27(2).

[96] Syryca, Finn, et al. "Automated generation of echocardiography reports using artificial intelligence: a novel approach to streamlining cardiovascular diagnostics." The International Journal of Cardiovascular Imaging. 2025; 1-11.

[97] Quer, Giorgio, and Eric J. Topol. "The potential for large language models to transform cardiovascular medicine." The Lancet Digital Health. 2024;6(10): e767-e771.

[98] Gendler, Moran, et al. "Large language models in cardiology: a systematic review." MedRxiv. 2024: 2024-09.