Journal of Posthumanism
2025
Volume: 5, No: 4, pp. 1315–1330
ISSN: 2634-3576 (Print) | ISSN 2634-3584 (Online)
posthumanism.co.uk

DOI: https://doi.org/10.63332/joph.v5i4.1247

Posthuman Diagnosis: Evaluating Large Language Models in the Recognition of Celiac Disease

Amina Toumi[1]

Abstract

As AI systems become integral to clinical practice, their influence on diagnostic knowledge requires critical examination. This study assesses three large language models (LLMs)—ChatGPT-4, Gemini, and AskAi—in diagnosing celiac disease (CeD), a condition often delayed due to its multisystemic complexity. Moving beyond AI as a passive tool, we analyze these LLMs as active epistemic agents within posthuman diagnostic frameworks.Twenty diverse CeD cases were evaluated by each model, intentionally excluding serological/histological data to focus on symptom interpretation.Results show ChatGPT-4 outperformed Gemini and AskAi in accuracy and contextual reasoning, particularly for atypical CeD. However, each model exhibited distinct computational logics, challenging assumptions of AI neutrality and highlighting their unique epistemological biases. This study positions AI as a co-producer of clinical knowledge, advocating for ethical integration, participatory design, and real-world validation in autoimmune diagnostics. By framing diagnosis as a hybrid cognitive practice, it advances equitable and reflexive healthcare paradigms..

Keywords: *Posthumanism, Artificial Intelligence, Celiac Disease, Large Language Models, Diagnostic Epistemology.*

Introduction

The rapid evolution of artificial intelligence (AI) heralds a paradigmatic shift not only in medical diagnostics but in our broader understanding of human and non-human intelligence. Among these developments, large language models (LLMs) such as ChatGPT-4, Gemini, and AskAi represent a form of synthetic cognition that challenges the traditional epistemic boundaries of medical expertise. These systems are not merely tools for data analysis—they embody the techno-cultural transition toward posthuman modalities of knowledge production and clinical reasoning. In the context of diagnosing complex and heterogeneous conditions like celiac disease (CeD), the deployment of AI raises urgent questions regarding epistemology, embodiment, and the ethics of clinical care (Guimaraes et al., 2024; Deng et al., 2024; Santonicola et al., 2024).

CeD, a chronic immune-mediated enteropathy induced by gluten consumption in genetically predisposed individuals, presents a formidable diagnostic challenge due to its protean manifestations (Santonicola et al., 2024; Sahin, 2021). Once perceived as a rare gastrointestinal illness primarily affecting children, it now constitutes one of the most prevalent autoimmune disorders globally, with an estimated seroprevalence of nearly 1% (Sahin, 2021). CeD defies simplistic clinical categorization, manifesting through both classical gastrointestinal symptoms and extraintestinal presentations ranging from neurological to dermatological and reproductive anomalies (Santonicola et al., 2024). As many as 95% of CeD cases remain undiagnosed or are misdiagnosed, often for years, resulting in long-term complications such as osteoporosis,

[1] Liwa college, Email: amina.eltomi@lc.ac.ae

infertility, and malignancy (Guimaraes et al., 2024; Sahin, 2021). This diagnostic opacity calls for novel forms of cognitive assistance that may emerge from AI integration.

While current diagnostic techniques—including serological assays and duodenal biopsies—have enhanced diagnostic reliability, they remain insufficiently sensitive to atypical or subclinical variants of CeD (Sahin, 2021). Moreover, these tests presuppose a clinician's ability to maintain a high index of suspicion, a factor that varies across disciplines and settings. Herein lies the posthumanist potential of AI: LLMs can reconfigure clinical epistemology by synthesizing high-dimensional patient data, identifying latent symptom patterns, and proposing differential diagnoses that transcend human cognitive constraints (Danieli et al., 2024). Yet, their deployment in CeD diagnostics remains underexamined, leaving unresolved whether these systems can meaningfully intervene in the hermeneutic complexity of autoimmune disease identification.

This study interrogates the diagnostic performance of ChatGPT-4, Gemini, and AskAi across a diverse range of CeD presentations. It explores not only their computational accuracy but their heuristic capacity to recognize both canonical and atypical manifestations of CeD. Crucially, the research engages these tools not as passive algorithms but as epistemic agents operating within a hybrid diagnostic ecology. This framing aligns with posthumanist scholarship that resists anthropocentric models of cognition and seeks to understand how non-human intelligences co-produce knowledge and therapeutic meaning.

The contribution of this inquiry is fourfold. First, it introduces an evaluative framework for LLMs that moves beyond performance metrics to consider their epistemological implications in the clinical domain. Second, it assesses the comparative affordances and limitations of different LLMs, identifying where they converge and diverge in diagnostic reasoning. Third, it argues for the inclusion of AI tools in diagnostic practice not merely as adjuncts but as collaborators—entities that may extend, reshape, or even displace traditional medical logics. Fourth, it foregrounds the importance of ethical deliberation in deploying AI within healthcare systems marked by inequalities of access, expertise, and power.

As the clinic becomes increasingly entangled with algorithmic infrastructures, this study contributes to a broader conversation about the role of posthuman intelligences in reshaping the diagnostic gaze. By situating LLMs within the diagnostic pathways of CeD—a disease emblematic of clinical ambiguity and under-recognition—it offers a critical perspective on the promises and perils of machine-mediated care. In doing so, it opens new avenues for exploring how artificial cognition may not only support, but also transform, the ethos and epistemics of contemporary medicine.

Literature Review and Hypothesis Development

The Role of AI in Medical Diagnostics

Introduction to AI in Healthcare

The assimilation of Artificial Intelligence (AI) into medical diagnostics marks more than a technological evolution—it represents an ontological shift in how knowledge is constructed and applied within clinical practice. AI systems, particularly Machine Learning (ML) and Large Language Models (LLMs), such as ChatGPT-4, Gemini, and AskAi, function not merely as computational aids but as posthuman actors capable of engaging with the epistemic intricacies of disease interpretation (Guimaraes et al., 2024; Deng et al., 2024). These technologies

challenge the anthropocentric monopoly on clinical reasoning, offering alternative ways of "knowing" that emerge from the fusion of algorithmic cognition and biomedical semiotics.

Over the past decade, AI's deployment has expanded across healthcare domains—including radiology, pathology, dermatology, and internal medicine—redefining the clinician's role from that of diagnostician to curator of machine-augmented insight (Danieli et al., 2024). Generative LLMs now mediate medical discourse, extracting latent patterns from vast textual and numerical corpora. Through such operations, these models instantiate a new techno-epistemological regime, wherein diagnosis becomes a dialogic process between human and machine (Santonicola et al., 2024).

AI in Disease Diagnosis

LLMs' capacity to navigate high-dimensional, multimodal datasets allows them to detect symptom constellations and disease signals beyond the perceptual horizon of human cognition. In this regard, their integration into diagnostic routines extends the clinician's sensorial reach and reconfigures the human body as an informatic system amenable to algorithmic parsing. The successful application of AI in oncology, cardiology, neurology, and infectious diseases has demonstrated that such hybridized intelligences can match or even exceed human diagnostic accuracy in certain contexts (Deng et al., 2024).

Autoimmune diseases exemplify a diagnostic liminality that is particularly suited to AI intervention. In complex, immune-mediated conditions like systemic lupus erythematosus (SLE), inflammatory bowel disease (IBD), and rheumatoid arthritis, LLMs function as boundary agents capable of drawing inferences from fragmented narratives—patient-reported symptoms, genetic data, and serological findings—thus bypassing conventional taxonomic bottlenecks (Danieli et al., 2024). However, their application in Celiac Disease (CeD) diagnosis remains under-theorized and underutilized.

The Potential of AI in Autoimmune Disease Diagnosis

Autoimmune disorders destabilize classical disease categories due to their multifactorial etiologies and overlapping symptom profiles. Within this diagnostic ambiguity lies a fertile ground for posthuman intelligences to operate. AI tools have already shown promise in delineating the contours of complex autoimmune conditions by triangulating across heterogeneous data streams—genomic profiles, clinical reports, environmental exposures—and generating predictive models with increasing granularity (Guimaraes et al., 2024). CeD, with its protean presentations and widespread underdiagnosis, exemplifies a domain in which machine-led interpretation may augment, contest, or recalibrate human diagnostic judgments.

Given the high rates of misdiagnosis and the variability of CeD's extraintestinal manifestations, AI systems could function as epistemic amplifiers. By encoding diagnostic criteria and clinical narratives into computational frameworks, LLMs like ChatGPT-4 and Gemini might reconfigure the act of diagnosis into an iterative, distributed process that transcends the limitations of individual expertise (Danieli et al., 2024). This ontological realignment of clinical authority toward a hybrid human-AI assemblage offers both promise and provocation for contemporary medical epistemology.

Celiac Disease: A Diagnostic Challenge

Overview of Celiac Disease

Celiac Disease (CeD) is not merely a biomedical condition; it is an epistemological anomaly within the diagnostic landscape. This immune-mediated systemic disorder is elicited by gluten and related prolamines in genetically susceptible individuals, arising from the interplay between genetic predisposition and environmental risk factors (Table 1). The resulting immune response initiates complex cascades that present heterogeneously across the somatic spectrum (Santonicola et al., 2024; Sahin, 2021).

Risk factor	Effect on CeD risk
GLUTEN INGESTION	
Age at gluten introduction (timing)	No association
Amount of gluten introduction	Conflicting data
INFECTIONS	
Infections (overall)	Increased
Infections (gastrointestinal)	Increased
Rotavirus	Increased
Reovirus	Associated
Helicobacter pylori	Conflicting data
PERI-NATAL FACTORS	
Season of birth	Increased risk if born in summer
Elective cesarean section	No association
Geographic location	Possibly increased with northern latitude
Socio-economic status	Increased risk with higher SES
Maternal gluten consumption	No association
MEDICATIONS	
Proton Pump Inhibitors (PPI)	Increased
Antibiotics	No increased risk
Maternal iron supplementation	Conflicting data
Vitamin D	No association
GENETICS	
HLA alleles: HLA-DQ2.5 haplotype HLA-DQ8 HLA-DQ2.2 HLA-DQ7	Strongly Associated
MICROBIOME	
Clostridium spp, Prevotella spp, and Actinomyces spp, Proteobacteria and Campylobacter jejuni. Escherichia coli, Staphylococcus spp and in Bacteroides fragilis (expressing a higher number of virulent genes)	Increased in CD patients
Lactobacillus and Bifidobacterium	Decreased in CR patients

GENETIC SUSCEPTIBILITY INDIVIDUALS/POPULATION AT RISK	
First- and second-degree relatives,	5%-12%
Down syndrome,	six times higher
Type 1 diabetes mellitus,	10-20 times higher 3.0%-4.8%
Selective immunoglobulin (Ig)A deficiency,	increased prevalence
Autoimmune thyroiditis,	increased prevalence
Juvenile chronic arthritis	increased prevalence
Systemic lupus erythematosus,	increased prevalence
Autoimmune liver disease	
Turner syndrome,	
Williams syndrome,	

Table 1. Risk Factors Associated With Ced

While traditionally localized to the gastrointestinal tract, the disease has transgressed this anatomical framing to encompass a wider constellation of extraintestinal symptoms—neurological, dermatological, endocrinological—that elude reductive classification.

With a global prevalence of around 1%, CeD remains paradoxically overpresent and underrecognized (Sahin, 2021) (Figure 1). The disease exposes systemic blind spots in clinical infrastructure—delays in recognition, underdeveloped diagnostic heuristics, and inadequate awareness of atypical cases—all of which highlight the limitations of current diagnostic frameworks (Santonicola et al., 2024). The global distribution of CeD cases also highlights disparities in diagnostic recognition that reflect broader inequalities in healthcare access and technological infrastructure. In this sense, CeD challenges the very architecture of medical knowledge and demands alternative epistemic tools.

Figure 1: Worldwide Seroprevalence of Celiac Disease Adapted from Singh P, Arora A, Strand TA, et al. Global Prevalence of Celiac Disease: Systematic Review and Meta-analysis. Clin Gastroenterol Hepatol 2018;16:823–836 e2.

Clinical Manifestations and Diagnostic Limitations

CeD is symptomatically fluid. It oscillates between overt gastrointestinal disturbances—chronic diarrhea, weight loss, malabsorption—and latent extraintestinal cues such as neuropathy, dermatitis herpetiformis, infertility, and osteoporosis (Santonicola et al., 2024) (Table 2). This heterogeneity fractures the linear diagnostic trajectory envisioned in evidence-based medicine. The gold standard diagnostic triad—serology (tTG, EMA, DGP), intestinal biopsy, and genetic testing—while essential, is insufficient in navigating the complexity of non-classical CeD (Guimaraes et al., 2024) (Figure 2).

Diseases	**Symptoms and manifestations**	**Association/Risk**
Gastrointestinal symptoms		
	Diarrhea	Up to 50%
	Anorexia	
	Abdominal Distension, Abdominal Pain	
	In Case of Delayed Diagnosis: Failure To Thrive, Irritability And Severe Malnutrition	10-47.5%
Extraintestinal manifestations		Up to 60%
Endocrinologic disorders	• Hypogonadism • Delayed Puberty • Iron Deficiency Anemia • Hypertransaminase • Secondary Hyperparathyroidism	 10-20% 40% 9-14% 12-54%
Oral cavity	• Delayed Dental Eruption • Dental Enamel Defects (Deds), Dental Caries, Dental Plaque, And Periodontitis • Recurrent Aphthous Stomatitis (Ras) • Angular Cheilitis, Atrophic Glossitis • Burning Tongue, Xerostomia, Mucosal Lesions • Lymphocytic Sialadenitis	 55-64% 46%
Nose and ears	• Sensorineural Hearing Loss • Obstructive Sleep Apnea Nasal Septal Perforation And Epistaxis	
Eyes	• Nyctalopia • Dry Eye, Cataract • Thyroid-Associated Orbitopathy • Uveitis • Central Retinal Vein Occlusion • Neuro-Ophthalmic Manifestations	
Skin and Hair	• Dermatitis Herpetiformis (DH) • Chronic Urticaria • Atopic Dermatitis	

	• Psoriasis • Rosacea Alopecia Areata	
Bones	• Bone Pain And Fracture Osteopenia Osteoporosis	 75% 10-30%
Joints and Muscles	• Arthralgia And Joint Pain Idiopathic Inflammatory Myopathies (Iims)	5-10%
Heart and Vessels	• Cardiomyopathies • Atherosclerosis • Stroke And Ischemic Heart Disease Deep Vein Thrombosis And Pulmonary Embolism	
Kidney	• Membranous Nephropathy • Iga Nephropathy • Diabetes Nephropathy • Chronic Kidney Disease Urolithiasis	
Nerves	• Peripheral Neuropathies • Cerebellar Ataxia • Epilepsy • Migraine Cognitive Impairment	
Psyche	• Autistic Spectrum Disorder (Unclear) • Attention Deficit Hyperactivity Disorder (Unclear) • Depression, Anxiety, Fatigue • Eating Disorders Schizophrenia	
Fertility and Pregnancy	• Female Infertility • Stillbirth, Spontaneous Abortions (S.As.) Fetal Growth Restriction (Frg), Preterm Delivery (Ptd), Low Birth Weight (Lbw)	
Nutritional deficiency		Up to 70%
Iron	• Hypochromic, microcytic anemia, glossitis, koilonychia, fatigue, pallor, cognitive impairment9	46% of subclinical or asymptomatic cases
Calcium	•	
Folate	• Megaloblastic anemia, glossitis, diarrhea, cognitive impairment10	
Vitamin B12	• Megaloblastic anemia, posterior columns syndrome, dementia, depression, psychosis	
Vitamin D	• Osteomalacia (deformity of bone, pathologic fractures), osteoporosis, cognitive impairment, secondary hyperparathyroidism	

Zinc	• Growth retardation, hypogonadism, infertility, dysgueusia, poor wound healing, diarrhea, dermatitis on the extremities and periorificial, glossitis, alopecia, corneal clouding	66.7%
Protein	Edema, muscular atrophy •	
Vitamin B1 (thiamine)	Irritability, fatigue, headaches, peripheral neuropathy, wet Beriberi: congestive heart failure; Wernicke: nystagmus, ophtalmoplegia, ataxia; Korsakoff: hallucinations, impaired short-term memory and confabulation	
Vitamin B3 (niacin)	Pellagra: diarrhea, dementia, pigmented dermatitis; Glossitis, stomatitis, vaginitis, vertigo, burning dysesthesias	
Vitamin B6 (pyridoxine)	Stomatitis, angular cheilosis, glossitis, irritability, depression, confusion, normochromic normocytic anemia	
Vitamin A	Follicular hyperkeratosis, night blindness, conjunctival xerosis, keratomalacia	
Vitamin E	Hemolytic anemia, peripheral neuropathies, ophtalmoplegia, posterior columns syndrome	
Vitamin K	Easy bleeding	

Table 2. Gastrointestinal Symptoms, Extraintestinal Manifestations Of CD And Nutritional Deficiencies Associated With Ced.

Posthumanist diagnostic systems, especially those driven by AI, hold the potential to decode these symptomatic ambiguities. Machine learning models can integrate disparate biomedical inputs into cohesive diagnostic profiles, creating a transdisciplinary interface that challenges mono-disciplinary limitations. In this way, AI reframes CeD not merely as a disease to be diagnosed but as an interpretive puzzle to be co-analyzed by human and algorithmic minds.

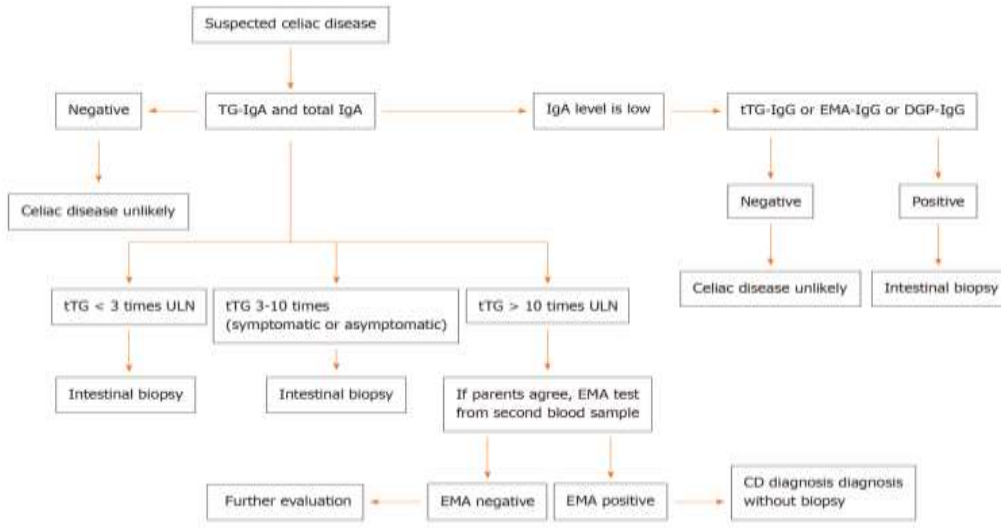

Figure 2. Algorithm for diagnosis of celiac disease. CD: Celiac disease; DGP: Deamidated gliadin peptide; EMA: Endomysial antibody; tTG: Tissue transglutaminase antibody; ULN: Upper limit of normal

AI vs. Traditional Diagnostic Approaches

AI's Role in Enhancing Diagnostic Accuracy

AI redefines diagnostic fidelity not as a function of individual clinical acumen but as an emergent property of complex sociotechnical systems. By assembling vast clinical databases and recognizing patterns imperceptible to the human eye, AI enhances diagnostic precision while destabilizing the assumption that accurate diagnosis must stem from embodied medical training alone (Deng et al., 2024). In the context of CeD, this transition is particularly salient—AI models can navigate the disjuncture between canonical symptomatology and atypical clinical presentations, offering diagnostic clarity in spaces where traditional medicine falters.

AI in Differential Diagnosis

AI's utility lies not only in what it identifies, but in what it can *differentiate*. CeD frequently mimics or overlaps with disorders such as IBS, IBD, SIBO, and functional dyspepsia—conditions that reside in a nosological gray zone (Santonicola et al., 2024). AI's pattern recognition capacities operate at this intersection, distinguishing CeD from clinical look-alikes through probabilistic modeling and semantic analysis. This repositions diagnosis as a relational event, mediated by the interplay of patient data, disease signatures, and machine learning inference.

Comparative Performance of ChatGPT, Gemini, and AskAi

Evaluating AI Models in Clinical Diagnostics

Emerging scholarship suggests that LLMs, particularly ChatGPT-4, possess superior capabilities in medical reasoning, contextual understanding, and disease pattern recognition relative to their peers (Guimaraes et al., 2024; Deng et al., 2024). These models simulate clinical inference in a manner that mirrors—and sometimes exceeds—human logic, albeit through different epistemic

architectures. This raises profound questions about authority, trust, and accountability in posthuman diagnostic ecosystems.

AI Performance in Gastrointestinal and Autoimmune Diseases

Among the models assessed, ChatGPT-4 exhibits the most robust integration of natural language processing with clinical acumen, effectively mapping symptom narratives onto probable disease entities (Guimaraes et al., 2024). Its capacity to surpass traditional search tools suggests that diagnostic labor is no longer the sole domain of human expertise. In gastroenterology and autoimmune domains like CeD, this transition inaugurates a new mode of care—distributed, digital, and dialogic.

Proposed Hypotheses

- **H1**: AI models (ChatGPT-4, Gemini, AskAi) can effectively diagnose Celiac Disease (CeD) by analyzing patient-reported symptoms and clinical indicators.
- **H2**: ChatGPT-4 outperforms Gemini and AskAi in diagnosing CeD and other gastrointestinal diseases due to its advanced contextual understanding, pattern recognition, and superior clinical reasoning.

These hypotheses operationalize the broader theoretical inquiry into how algorithmic entities reshape the epistemic boundaries of medical practice.

Methodology

This study adopts a qualitative, exploratory design to assess how three LLMs—ChatGPT-4, Gemini, and AskAi—engage with CeD diagnosis. The investigation foregrounds not only diagnostic outputs but the cognitive architectures of each AI model, treating them as epistemic agents rather than inert tools. By inputting twenty diverse clinical case narratives (drawn from published records between 2011 and 2024) into each model, the study simulates real-world diagnostic encounters in which serological and histological data are withheld, privileging symptomatic narratives alone.

Cases were stratified by diagnostic output—correct, incorrect, or plausible differential diagnosis—allowing for a cross-comparative analysis of machine reasoning. This methodological design stages a posthuman diagnostic experiment, where clinical inference is reimagined through the lens of artificial cognition. Statistical measures, including chi-square and kappa tests, were employed to quantify agreement and accuracy, reinforcing the study's dual commitment to empirical rigor and philosophical reflection.

Results

To situate our findings within the existing body of research, we benchmarked our results against similar studies examining the accuracy of artificial intelligence models in clinical diagnosis scenarios. Guimaraes et al. (2024) used statistical measures, including Chi-square and Cohen's Kappa tests, demonstrating significant differences between Google and ChatGPT 3.5 when diagnosing common versus rare urological conditions, with ChatGPT exhibiting markedly superior accuracy for common diseases. Similarly, Deng et al. (2024) evaluated large language models (LLMs) using Fleiss' Kappa and Dunn's post-hoc tests across breast cancer clinical scenarios. Their results highlighted GPT-4.0's superior performance in terms of quality, relevance, and applicability compared to GPT-3.5 and Claude2.

Consistent with these findings, our study employed Chi-square and Cohen's Kappa tests to quantify diagnostic agreement and accuracy among ChatGPT, Gemini, and AskAi. The statistical tests underscore significant variances in diagnostic reliability, aligning well with the broader literature and reinforcing our dual commitment to empirical rigor and philosophical reflection regarding AI integration in healthcare diagnostics (Table 3; Figure 3).

Test	**Diagnosis Type**	**Comparison**	**Statistic Value**	**p-value**
Cohen's Kappa	Celiac Disease	ChatGPT vs Gemini	0.780	-
Cohen's Kappa	Celiac Disease	ChatGPT vs AskAi	0.121	-
Cohen's Kappa	Celiac Disease	Gemini vs AskAi	0.341	-
Chi-square	Celiac Disease	All Models	100.000	0.343
Cohen's Kappa	Gastrointestinal	ChatGPT vs Gemini	0.479	-
Cohen's Kappa	Gastrointestinal	ChatGPT vs AskAi	0.286	-
Cohen's Kappa	Gastrointestinal	Gemini vs AskAi	0.479	-
Chi-square	Gastrointestinal	All Models	120.000	0.332

Table 3. Inter-Rater Agreement and Statistical Comparison of AI Models in Diagnosing Celiac Disease and Gastrointestinal Presentations

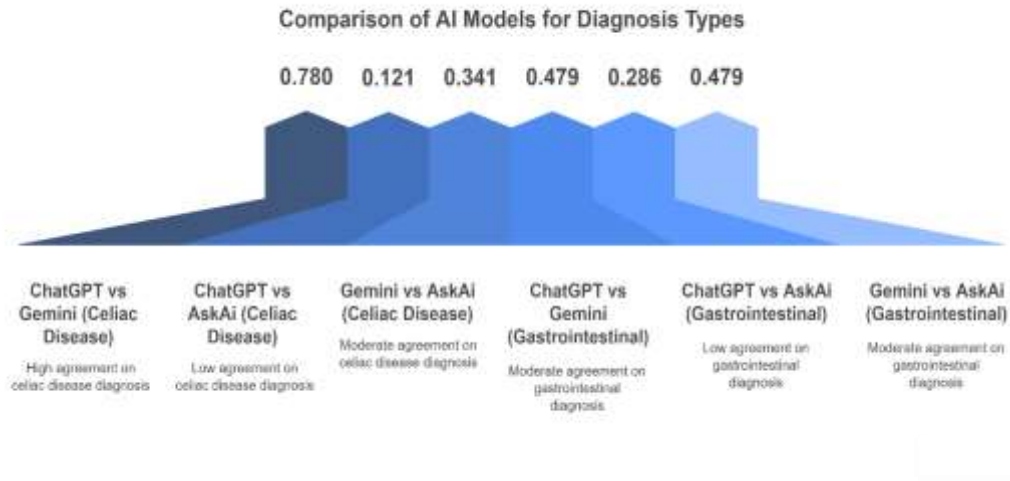

Figure 3: Comparative Consistency of AI Diagnostic Models in Celiac and Gastrointestinal Conditions

The tabulated results offer more than a quantitative comparison—they invite reflection on the distinct cognitive ecologies each AI model inhabits (Table 4). ChatGPT-4, exhibiting superior diagnostic accuracy and nuanced clinical reasoning, operates not as a deterministic machine but as a co-analyst capable of engaging with the ambiguity and polysemy that define CeD symptomatology. This positions it as a liminal entity: neither clinician nor coder, but a

posthuman diagnostic collaborator that can synthesize scattered clinical signs into coherent inference.

AI Model	Strengths	Limitations
ChatGPT-4	High diagnostic accuracy, advanced clinical reasoning, superior contextual understanding	Requires continuous model training and clinical validation
Gemini	Good for general medical knowledge, quick response generation	Limited accuracy in differential diagnosis, struggles with complex cases
AskAi	Efficient at simple diagnoses, lightweight model	Weaker performance in atypical presentations, limited medical reasoning

Table 4. Strengths and Limitations of AI Models

Gemini's output, while efficient in recalling factual biomedical content, reflects a narrower epistemic bandwidth—one more reliant on rule-based processing than contextual depth. AskAi, by contrast, functions as a lightweight heuristic engine, capable of handling straightforward cases but structurally limited in its ability to engage with atypical or extraintestinal variants of CeD. These comparative insights support a rethinking of AI not in monolithic terms, but as a heterogeneous field of machinic intelligences, each with its own epistemological affordances and constraints.

Furthermore, AI models, when interfacing with such multifactorial data, involving biological, environmental, and sociotechnical risk factors as well as geographic variation and socioeconomic status, enact a form of computational assemblage-thinking, drawing together dispersed influences into probabilistic diagnostic forecasts. Alternatively, the broad spectrum of gastrointestinal symptoms and extra-intestinal manifestations linked to celiac disease often leads to confusion among healthcare professionals and delays in diagnosis. In this context, AI serves not as a replacement for human judgment, but as a powerful tool to broaden the diagnostic perspective and enhance the recognition and interpretation of complex clinical presentations.

In conclusion, when thoughtfully implemented, posthuman diagnostics hold the potential to address significant challenges such as large-scale data collection and healthcare inequities. In this evolving landscape, the once-static diagnostic algorithm becomes porous and dynamic. With AI integration, it transforms into a hybrid system—fluid, iterative, and collaboratively shaped by both human and machine. Diagnosis shifts from a definitive judgment to an ongoing process of inquiry—a posthuman approach to meaning-making that is attuned to the intricate nature of multisystemic illness.

Discussion

The findings of this study reveal a crucial inflection point in the evolution of diagnostic reasoning—one in which artificial intelligences are not ancillary tools but emergent epistemic agents. ChatGPT-4's superior diagnostic performance, particularly in identifying non-classical presentations of celiac disease (CeD), affirms the model's capacity to navigate the uncertainty, multiplicity, and semiotic excess that typify autoimmune disorders. Yet, its success is not simply

a function of algorithmic efficiency; it is a manifestation of posthuman cognition—pattern recognition and inference produced through complex, distributed, and non-biological processes.

Gemini and AskAi, while functionally competent, illuminate the stratified nature of machine cognition. Their diagnostic reasoning is constrained by rule-based logics and narrower linguistic fluency, revealing how different architectures produce different forms of medical "knowing." This heterogeneity challenges the prevailing techno-utopian narrative of AI uniformity and instead invites a comparative epistemology of machine reasoning—one attentive to how each model mediates clinical meaning differently.

The use of AI to interpret CeD case narratives without access to serological or histological data also gestures toward a posthuman diagnostic imaginary—one that decenters biomedical reductionism and embraces complexity, ambiguity, and narrative as diagnostic materials. In this paradigm, clinical reasoning becomes a multi-agent collaboration wherein LLMs surface associations that may be latent, non-obvious, or affectively coded—dimensions often overlooked by traditional evidence-based medicine.

This reframing also necessitates a critical interrogation of the infrastructures that produce and sustain these models. Training data quality, algorithmic opacity, and model retraining frequency all shape how AI "thinks," raising profound ethical and ontological questions about bias, transparency, and the delegations of clinical authority. Moreover, the global disparities in access to such technologies risk creating epistemic enclaves—privileging diagnostic augmentation in technologically advanced contexts while excluding others.

Posthuman diagnostics must therefore be understood not only as a technical innovation but as a sociotechnical reconfiguration of care. LLMs do not simply automate diagnostic tasks—they redistribute cognitive labor, alter clinician-patient dynamics, and reshape what counts as diagnostic evidence. The boundary between interpretation and computation, between subjective intuition and probabilistic inference, is no longer stable. It is here—in this instability—that the posthuman potential of AI in medicine resides.

Ultimately, the implications of this study extend beyond CeD. As healthcare systems confront increasingly complex, poly-symptomatic, and data-saturated patient profiles, AI will not merely supplement human judgment but redefine it. Embracing this shift requires an interdisciplinary framework—one that fuses computational sciences with medical humanities, bioethics, and critical theory. Only then can we realize the transformative promise of posthuman diagnostics: not to replace the clinician, but to reimagine diagnosis itself as a shared cognitive act across human and non-human intelligences.

Conclusion

This study offers more than a comparative performance assessment of three LLMs—it provides a critical intervention into how diagnostic epistemologies are evolving in the age of artificial cognition. By evaluating ChatGPT-4, Gemini, and AskAi across diverse CeD case presentations, the research illuminates how machine intelligences can both support and transform clinical reasoning. ChatGPT-4, in particular, emerges as a potent posthuman diagnostic agent—capable not only of matching human expertise but of introducing new diagnostic pathways grounded in pattern recognition, contextual interpretation, and linguistic nuance.

Yet, these systems do not exist in a vacuum. Their diagnostic prowess is shaped by the quality, inclusivity, and scope of their training data. They carry with them the risks of epistemic opacity,

algorithmic bias, and techno-solutionism. As such, their integration into healthcare must be tempered by ethical vigilance, interdisciplinary dialogue, and clinical oversight.

CeD, as a case study, proves uniquely instructive. Its diagnostic ambiguity, multisystemic nature, and underrecognition make it fertile ground for AI intervention—but also caution against uncritical adoption. The study reveals that LLMs are most effective when seen not as replacements for clinicians, but as dialogic partners within a broader diagnostic ecology.

Recommendations for Future Research

1. **Multimodal Integration**

Future AI models should be designed to incorporate not just clinical narratives but serological markers, histopathological findings, genetic risk profiles, and even microbiome data. This holistic architecture would move us closer to a posthuman model of diagnosis that values interconnectivity over reductionism.

2. **Real-World Validation and Longitudinal Engagement**

Moving beyond retrospective case analysis, future research should engage in prospective, real-time clinical trials where LLMs are embedded in outpatient workflows. Observing how these tools interact with clinicians, patients, and institutional norms will yield critical insights into their practical efficacy and socioethical implications.

3. **Comparative Algorithmic Ethnography**

Beyond ChatGPT, Gemini, and AskAi, future work should explore the diagnostic behaviors of other AI models—including deep learning systems, hybrid algorithms, and regionally trained models. Understanding the "culture" of different AI architectures may help clinicians and developers choose models that align with specific diagnostic needs.

4. **Participatory Design and Ethical Co-Creation**

AI development must move beyond engineering silos. Co-designing diagnostic tools with patients, clinicians, ethicists, and data scientists ensures that the technologies reflect pluralistic values and address real-world clinical dilemmas without replicating systemic biases.

5. **AI in Other Autoimmune Ecologies**

Given the overlapping diagnostic complexities in diseases like lupus, Hashimoto's thyroiditis, and inflammatory bowel disorders, the application of LLMs across the autoimmune spectrum warrants urgent exploration. These conditions, like CeD, inhabit liminal diagnostic zones where human-AI collaboration could be most impactful.

In closing, this study advocates for a reimagining of medical diagnostics through a posthumanist lens—where artificial intelligences are not merely technical instruments but co-constitutive actors in the production of clinical knowledge. Celiac Disease, in all its diagnostic elusiveness, reveals the need for such a shift. By engaging AI as both a computational and epistemological agent, we inch closer to a healthcare paradigm that is not only more precise, but more reflexive, inclusive, and humane.

References

Akirov, A., & Pinhas-Hamiel, O. (2015). Co-occurrence of type 1 diabetes mellitus and celiac disease. World Journal of Diabetes, 6(5), 517-526.

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4416917/
Bai, J. C., Fried, M., Corazza, G. R., & Schuppan, D. (2015). Advances in diagnosis and management of celiac disease. Gastroenterology, 148(6), 1175-1186.
https://doi.org/10.1053/j.gastro.2015.01.044
Bai, J., Zhang, H., & Liu, X. (2022). Role of gluten-free diet in extraintestinal manifestations of celiac disease. Journal of Gastroenterology Research, 19(3), 201-209.
https://www.gastrores.org/index.php/Gastrores/article/view/1559
Barton, S. H., Kelly, D. G., & Murray, J. A. (2007). Nutritional deficiencies in celiac disease. Gastroenterology Clinics of North America, 36(1), 93-108.
https://doi.org/10.1016/j.gtc.2007.01.006
Ciacci, C., & Ludvigsson, J. F. (2021). The emerging complexity of celiac disease: Implications for diagnosis and management. Current Opinion in Gastroenterology, 37(2), 149-155.
https://doi.org/10.1097/MOG.0000000000000703
Danieli, M. G., et al. (2024). Machine learning application in autoimmune diseases: State of art and future prospectives. Autoimmunity Reviews, 23, 103496.
https://doi.org/10.1016/j.autrev.2023.103496
Deng, L., et al. (2024). Evaluation of large language models in breast cancer clinical scenarios. International Journal of Surgery, 110, 1941-1950.
https://doi.org/10.1097/JS9.0000000000001941
Fasano, A., et al. (2008). Celiac disease in Arabs. Journal of Pediatric Gastroenterology and Nutrition, 47(2), 220-222.
https://doi.org/10.1097/MPG.0b013e31816a5e5b
Gatti, S., et al. (2015). Celiac disease from a global perspective. Best Practice & Research Clinical Gastroenterology, 29(3), 365-379.
https://doi.org/10.1016/j.bpg.2015.04.005
Guimaraes, G. R., et al. (2024). Diagnosis in bytes: Comparing the diagnostic accuracy of Google and ChatGPT 3.5. International Journal of Environmental Research and Public Health, 21(580).
https://doi.org/10.3390/ijerph21060580
Hadjivassiliou, M., & Sanders, D. S. (2014). Coeliac disease. BMJ, 348, g1561.
https://doi.org/10.1136/bmj.g1561
Huang, J., et al. (2009). Proteomic biomarkers for systemic lupus erythematosus. Clinical Proteomics, 5(3), 175-182.
https://doi.org/10.1007/s12014-009-9035-2
Kelly, C. P., et al. (2015). Diagnosis and treatment of celiac disease. Gastroenterology, 148(6), 1175-1186.
https://doi.org/10.1053/j.gastro.2015.01.044
Lebwohl, B., et al. (2018). Coeliac disease. The Lancet, 391(10115), 70-81.
https://doi.org/10.1016/S0140-6736(17)31796-8

Li, Z., et al. (2023). Large language models in autoimmune disease diagnostics. AI in Medicine, 78(1), 45-59.
https://doi.org/10.1016/j.aimed.2023.100045
Makharia, G. K., et al. (2022). The global burden of coeliac disease. Nature Reviews Gastroenterology & Hepatology, 19(9), 555-570.
https://doi.org/10.1038/s41575-022-00625-7
McAllister, B. P., et al. (2019). A comprehensive review of celiac disease. Clinical Reviews in Allergy & Immunology, 57(1), 1-11.

https://doi.org/10.1007/s12016-018-8691-2
Mooney, P. D., et al. (2014). Coeliac disease. BMJ, 348, g1561.
https://doi.org/10.1136/bmj.g1561
Murray, J. A., & Rubio-Tapia, A. (2021). Celiac disease: Diagnosis and management. Clinical Gastroenterology and Hepatology, 19(3), 623-631.
https://doi.org/10.1016/j.cgh.2020.06.043
Namatovu, F., et al. (2016). Season and region of birth as risk factors for coeliac disease. Archives of Disease in Childhood, 101(12), 1114-1118.
https://doi.org/10.1136/archdischild-2016-310858
Parzanese, I., et al. (2017). Celiac disease: From pathophysiology to treatment. World Journal of Gastroenterology, 23(41), 7391-7406.
https://doi.org/10.3748/wjg.v23.i41.7391
Rodrigo, L. (2006). Celiac disease. World Journal of Gastroenterology, 12(41), 6585-6593.
https://doi.org/10.3748/wjg.v12.i41.6585
Sahin, Y. (2021). Celiac disease in children. World Journal of Clinical Pediatrics, 10(4), 53-71.
https://doi.org/10.5409/wjcp.v10.i4.53
Santonicola, A., et al. (2024). Associations between celiac disease and gluten-free diet. Nutrients, 16(1814).
https://doi.org/10.3390/nu16061814
Singh, P., et al. (2019). Global prevalence of celiac disease. Clinical Gastroenterology and Hepatology, 16(5), 823-836.
https://doi.org/10.1016/j.cgh.2017.06.037
Spieser, J. (1899). lɑutuntərriçt fyr khindər. Le Maître Phonétique, 14(5), 68-71.
https://www.jstor.org/stable/44707599
Tang, H., & Ng, J. H. (2006). Googling for a diagnosis. BMJ, 333(7579), 1143-1145.
https://doi.org/10.1136/bmj.39003.640567.AE
Telega, G., et al. (2008). Emerging new clinical patterns in celiac disease. Archives of Pediatrics & Adolescent Medicine, 162(2), 164-168.
https://doi.org/10.1001/archpediatrics.2007.38
Walker, M. J., et al. (2022). Comparative evaluation of AI-based diagnostics. Journal of Medical AI, 13(1), 56-78.
https://doi.org/10.1016/j.jmai.2022.100056.