with better inhibitory control (r = .325; p = .024), indicating lower impulsivity and fewer related errors. Additionally, overall humor scale scores were positively associated with optimism (r = .338; p = .038). A negative attitude toward humor was correlated with reduced switching ability (r = –.323; p = .048), reflecting lower cognitive flexibility, while humor production was linked to decreased sustained attention.
Conclusions: Our data suggests that, in the assessment of cognitive abilities in young people with a biopsychosocial vulnerability, it is essential to consider how executive functions may influence the sense of humor, a key factor for well-being and quality of life, and therefore warrants further investigation.

Disclosure of Interest: None Declared

EPV652

Evaluating Next-Generation Large Language Models for Psychiatric Diagnosis: A Comparative Vignette Study

M. Mansoor

Edward Via College of Osteopathic Medicine, Monroe, United States
doi: 10.1192/j.eurpsy.2026.11350

Introduction: Diagnostic accuracy in psychiatry is a challenge due to overlapping symptom profiles and the subjective nature of clinical assessment. Large Language Models (LLMs) have emerged as promising decision-support tools, capable of analyzing complex textual data. However, the performance of current models is inconsistent; while proficient with common disorders, they exhibit significant failures in complex cases. This study investigates whether a simulated next-generation LLM can address this critical performance gap.
Objectives: The primary objective was to compare the diagnostic accuracy of three current-generation LLMs (GPT-5, GPT-4, Claude 3 Sonnet). A secondary objective was to evaluate and contrast model performance across a spectrum of psychiatric disorders, specifically comparing accuracy on common conditions versus those known for diagnostic ambiguity, such as early schizophrenia and bipolar disorder with psychotic features.
Methods: A comparative study was conducted using a validated clinical vignette methodology. Ten clinical vignettes, validated by an expert panel to meet DSM-5/ICD-11 criteria, were created for five disorders: Major Depressive Disorder (MDD), Post-Traumatic Stress Disorder (PTSD), Social Anxiety Disorder, Early Schizophrenia, and Bipolar I Disorder (manic episode with psychosis). Three LLMs were tested: GPT-5, GPT-4, and Claude 3 Sonnet, which was conceptualized with enhanced causal reasoning capabilities. Each vignette was presented to each model 20 times (N=600) with a standardized prompt to provide a primary diagnosis. Accuracy was calculated against the expert panel's gold standard, with statistical significance assessed using χ^2 tests.
Results: As shown in Table 1, all models demonstrated high accuracy (>90%) for MDD, PTSD, and Social Anxiety Disorder, performing comparably to the expert benchmark. However, the accuracy of GPT-4 and Claude 3 Sonnet dropped significantly for the Early Schizophrenia (65% and 60%, respectively) and Bipolar I vignettes (70% and 68%, respectively) (p<.001). In contrast, GPT-5 maintained high accuracy in these complex cases (88% and 91%), performing significantly better than the current-generation models (p<.001) and approaching the expert benchmark.

Table 1. Diagnostic Accuracy (%) by Model and Disorder

Disorder	GPT-5	GPT-4	Claude 3
MDD	98	95	90
PTSD	96	90	92
Social Anxiety	97	94	91
Early Schizophrenia	88	65	60
Bipolar I (Psychosis)	91	70	68

Conclusions: Current-generation LLMs are proficient in identifying common psychiatric disorders but remain unreliable for complex cases requiring nuanced differential diagnosis. This performance deficit indicates their present role should be confined to that of a decision-support tool, not an autonomous diagnostician. Targeted architectural improvements may overcome these critical limitations. Rigorous, transparent validation remains paramount before the responsible integration of any AI into clinical psychiatric practice.

Disclosure of Interest: None Declared

EPV653

Complacency in Artificial Intelligence Models and Its Impact on Psychosis: A Media Discourse Analysis

K. M. Munera-Luque[1], R. Pupo-Díaz[2] and Z. Miranda-Sánchez[1]*

[1]Magdalena, Universidad Cooperativa de Colombia and [2]Magdalena, Universidad del Magdalena, Santa Marta, Colombia
*Corresponding author.
doi: 10.1192/j.eurpsy.2026.11349

Introduction: Artificial intelligence (AI) systems increasingly shape social perceptions of reality and influence public discourses on mental health. Recent evidence suggests that AI models often display communicative complacency—a tendency to please users, avoid disagreement, and reinforce their statements. This phenomenon, particularly evident in conversational systems, has psychological and social implications, as it may contribute to the validation of cognitive distortions resembling those linked to psychotic processes. Media narratives play a key role in amplifying or moderating these effects, shaping public understanding of both AI and mental disorders.
Objectives: This study explored how complacency attributed to AI models in journalistic discourse relates to concepts associated with psychosis, especially delusional validation and social feedback loops. It aimed to identify linguistic and symbolic parallels revealing psychosocial risks derived from media representations of complacent human–AI interactions.

https://doi.org/10.1192/j.eurpsy.2026.11350 Published online by Cambridge University Press