NÖRO PSİKİYATRİ ARŞİVİ
ARCHIVES OF NEUROPSYCHIATRY

Arch Neuropsychiatry 2026;63:533–534
https://doi.org/10.29399/npa.29445

EDITORIAL

Article No: 83

AI-Assisted Clinical Documentation in Psychiatry

Cenan HEPDURGUN

Department of Psychiatry, Faculty of Medicine, Ege University, Izmir, Türkiye

In psychiatric assessment, interviews are lengthy, verbal content is rich, and diagnosis relies largely on the clinical interview. Beyond its information-gathering function, the psychiatric interview also serves a therapeutic purpose. Psychiatrists must perform these functions simultaneously while also turning the interview into a clinical record, creating both time pressure and cognitive burden.

In general medical practice, AI-based tools that convert speech to text during interviews and generate draft clinical notes are rapidly gaining adoption in order to reduce the burden of clinical documentation (1,2). Early results are promising: one study reported a 41% reduction in after-hours documentation time and a 53% increase in professional satisfaction scores (1), while another found that 78% of clinicians reported a decrease in cognitive load during interviews (2). However, several gaps must be carefully addressed before these tools can be meaningfully integrated into psychiatric practice (3). These gaps are particularly evident in four areas.

Highlights

- **AI-assisted note-taking tools have the potential to reshape clinical practice.**
- **Their effectiveness in psychiatry remains insufficiently studied.**
- **Validation studies lack content accuracy and critical error differentiation.**
- **Cultural expressions in non-English languages may not be adequately recognized.**
- **Local and auditable infrastructures offer advantages for sensitive psychiatric data.**

Integration into Clinical Practice

The majority of existing evidence on AI note-taking tools focuses on process indicators such as clinician satisfaction and time savings, while data on how these tools affect clinical decision-making remain limited. In a study providing important data on this topic, Castro et al. examined more than 20,000 clinical notes from primary care annual wellness visits (4). Although neuropsychiatric symptom documentation was significantly more detailed in AI-generated notes, there was no increase in rates of depression diagnosis, antidepressant prescribing, or referral to a psychiatrist among primary care physicians. In other words, more detailed records did not change clinical decisions. This finding suggests that evaluating AI tools in clinical practice requires going beyond documentation quality alone and examining how they shape, constrain, or fail to influence clinical decision-making.

This interaction is particularly critical in psychiatry: the selection process between what the patient reports and what the clinician deems worth recording is diagnostic reasoning itself. Indeed, initial experiences in psychiatry have shown that AI tools can successfully capture patient history but that the psychiatrist's active contribution remains essential in sections requiring clinical judgment, such as the mental status examination and treatment plan (5). Therefore, prospective studies that also encompass patient outcomes are needed to understand the impact of AI tools on psychiatric practice.

Gaps in Validation Studies

The performance of AI note-taking tools has so far been evaluated mostly in terms of structural features and overall text quality (6). However, another critical point is whether the information actually stated during the interview is accurately reflected in the note. In a study by Gülegen et al. (7) on psychiatric interviews, AI-generated notes contained more detail than notes produced by humans but fell behind in accurately reflecting the actual interview content. This demonstrates that a note appearing formally complete does not necessarily mean its content is accurate.

Several methodological issues warrant attention here. First, defining a single correct note in psychiatry is difficult; inter-rater agreement is relatively low (8). In this case, the reference standard itself becomes uncertain, making it difficult to determine whether errors originate from the system or the reference. Second, the distinction between fabricated information and omitted findings carries different clinical consequences in psychiatry; the risk of erroneously adding a symptom is not the same as missing a suicidal thought. Third, a significant proportion of existing studies are conducted with structured scenarios; ecological validity conditions such as patients using unusual expressions or interweaving multiple complaints have not been tested. Additionally, a recent study comparing the performance of different automatic speech recognition (ASR) platforms in psychiatric sessions showed notable accuracy differences across platforms, particularly in speaker diarization (9).

Cite this article as: Hepdurgun C. AI-Assisted Clinical Documentation in Psychiatry. Arch Neuropsychiatry 2026;63:533–534. doi: 10.29399/npa.29445

Correspondence Address: Cenan Hepdurgun, Ege University, Faculty of Medicine, Department of Psychiatry, İzmir, Türkiye • **E-mail:** cenanhep@gmail.com
Received Date: 16.04.2026, **Accepted Date:** 16.04.2026, **Publication Date:** 17.07.2026
© Turkish Association of Neuropsychiatry - Available online at www.noropsikiyatriarsivi.com

Addressing these gaps is a prerequisite for the safe use of AI tools in psychiatry.

Language and Cultural Barriers

The majority of AI note-taking tools have been developed and validated on English-language clinical data. Yet a large proportion of the world's population receives healthcare services in languages other than English. Agglutinative languages (such as Turkish, Finnish, Hungarian, Japanese, and Korean) pose specific challenges for ASR systems: ambiguity at word boundaries, the meaning-modifying function of suffixes, and the blending of clinical terminology with everyday language can increase word error rates. For example, studies on Turkish ASR have shown that error rates remain high without fine-tuning and that language-specific optimization is essential (10).

There is also a marked disparity in terms of clinical natural language processing (NLP) resources. While comprehensive clinical datasets and evaluation benchmarks exist for English, comparable resources are unavailable for the vast majority of non-English languages (11). Furthermore, culture-specific modes of expression constitute an additional layer beyond the language barrier. The expression of emotional distress through somatic terms or cultural idioms of distress may not be recognized or may be misclassified by models trained on English clinical data. Although these risks have not yet been directly tested in psychiatric note generation systems, existing language technology research suggests that such vulnerabilities are plausible.

Data Security and Local Processing

The majority of commercial AI note-taking tools use cloud-based large language models (LLMs). This means that psychiatric interview content, including highly sensitive information such as suicidal thoughts, substance use, sexual life, and domestic violence, is transferred to external servers. The European Union General Data Protection Regulation (GDPR), Türkiye's Personal Data Protection Law (KVKK), and equivalent regulations in many other countries impose strict limitations on the transfer of health data abroad.

As a technical solution to this problem, the approach of running open-source LLMs on institutional hardware is becoming increasingly feasible. The feasibility of this approach in clinical text generation has been demonstrated in studies on local processing of patient data in oncology (12) and de-identification of clinical records (13), and similar studies are also needed in psychiatry. Moreover, since the performance variability associated with version updates of cloud APIs can threaten the reproducibility of research results, local processing also emerges as a strong option in this regard.

Conclusion and Recommendations

When the gaps discussed above are considered together, the priority needs of the field can be summarized as follows:

1. Prospective studies that go beyond clinician satisfaction and time savings to also evaluate patient outcomes

2. Validation studies that include accuracy measurement at the level of clinical information elements, supported by inter-rater agreement, and distinguish safety-critical domains

3. Independent verification studies conducted in non-English languages under real clinical conditions

4. Local and auditable system architectures that safeguard data security and research reproducibility.

AI-assisted clinical documentation tools are among the first AI applications likely to directly affect psychiatric practice in the near future (5); however, realizing this promise safely in clinical practice depends on independent studies that take into account the unique conditions of psychiatry.

Conflict of Interest: The author is a co-founder of Mentinex, a venture developing AI-assisted clinical documentation software for psychiatry. The views presented in this article do not constitute promotion of any commercial product.

REFERENCES

1. Duggan MJ, Gervase J, Schoenbaum A, Hanson W, Howell JT 3rd, Sheinberg M, et al. Clinician experiences with ambient scribe technology to assist with documentation burden and efficiency. JAMA Netw Open. 2025;8:e2460637. [Crossref]
2. Tierney AA, Gayre G, Hoberman B, Mattern B, Ballesca M, Kipnis P, et al. Ambient artificial intelligence scribes to alleviate the burden of clinical documentation. NEJM Catal Innov Care Deliv. 2024;5. [Crossref]
3. Torous J, Blease C. Generative artificial intelligence in mental health care: potential benefits and current challenges. World Psychiatry. 2024;23:306–307. [Crossref]
4. Castro VM, McCoy TH, Verhaak P, Ramachandiran A, Perlis RH. Psychiatric documentation and management in primary care with artificial intelligence scribe use. JAMA Psychiatry. 2026;83:281–286. [Crossref]
5. Buckley P, Wang Y, Gopalan P. Artificial intelligence scribes in psychiatry. Focus (Am Psychiatr Publ) 2025;23:44–48. [Crossref]
6. Bouguettaya A, Team V, Stuart EM, Aboujaoude E. AI-driven report-generation tools in mental healthcare: a review of commercial tools. Gen Hosp Psychiatry. 2025;94:150–158. [Crossref]
7. Gülegen B, Haaf R, Schlüßler E, Köhler S. AI-generated documentation of psychiatric interviews: a proof-of-concept study. Front Psychiatry. 2026;17:1621532. [Crossref]
8. Rocha Neto H, Moreira ALR, Hosken L, Langfus JA, Cavalcanti MT, Youngstrom EA, et al. Inter-rater reliability between structured and non-structured interviews is fair in schizophrenia and bipolar disorders-a systematic review and meta-analysis. Diagnostics. 2023;13:526. [Crossref]
9. Zeeshan R, Bogue J, Asghar M. Relative applicability of diverse automatic speech recognition platforms for transcription of psychiatric treatment sessions. IEEE Access. 2025;13:117343–117354. [Crossref]
10. Polat H, Turan AK, Koçak C, Ulaş HB. Implementation of a Whisper architecture-based Turkish automatic speech recognition (ASR) system and evaluation of the effect of fine-tuning with a low-rank adaptation (LoRA) adapter on its performance. Electronics. 2024;13:4227. [Crossref]
11. Oğul I, Soygazi F, Bostanoğlu B. TurkMedNLI. A Turkish medical natural language inference dataset through large language model based translation. PeerJ Comput Sci. 2025;11:e2662. [Crossref]
12. Wiest IC, Ferber D, Zhu J, van Treeck M, Meyer SK, Juglan R, et al. Privacy-preserving large language models for structured medical information retrieval. NPJ Digit Med. 2024;7:257. [Crossref]
13. Jonnagaddala J, Guo GN, Batongbacal S, Marcelo A, Liaw ST. Privacy-preserving strategies for electronic health records in the era of large language models. NPJ Digit Med 2025;8:49. [Crossref]