doximity

Resident Ratings of Clinical AI Tools

Executive Summary

This report summarizes results from 4,688 side-by-side evaluations in which physician residents compared responses from Doximity Ask with those from other clinical AI tools (e.g., OpenEvidence, UpToDate, ChatGPT) using clinical questions they personally posed on each platform.

Across evaluations, most questions produced clinically similar answers—64% marked the preferred clinical answer as "Nearly the same" or "Bit better". This is expected given their reliance on comparable foundational language models.

When a preference was expressed, Doximity Ask was favored at nearly three times the rate of the nearest competitor (69% vs 24%) when tallying which answer was best. Reasons cited for Doximity Ask included:

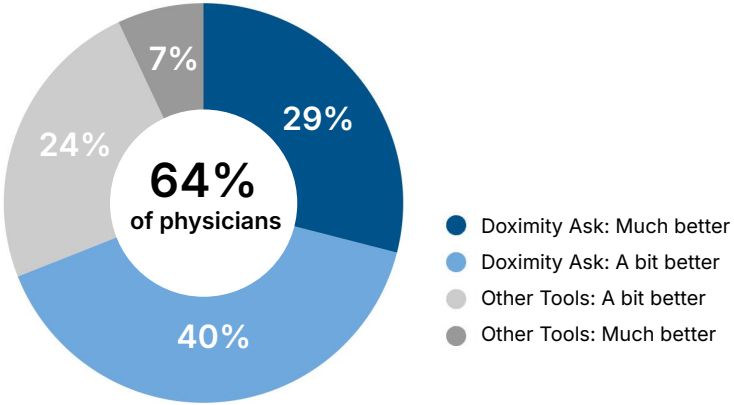
- Full PDF access to over 2,000 medical journals
- Built-in deterministic drug reference
- Absence of full-screen banner ads
- Tables and formatting that were easier to read
- Faster response times

Doximity Ask was most often preferred for complex or niche evidence queries, reflecting the value of direct access to full-text primary literature.

Similar Sources & Models

Most questions (64%) yielded answers that were “Nearly the same” or a “Bit better” across tools, a product of high correlation among sources and foundational models used.

Magnitude of difference across tools*



*”Bit Better” represents who selected either “Bit better” or “Nearly the same”, and “Much better” represents respondents who selected “Much better” or “Far better”

doximity

Overall Ratings

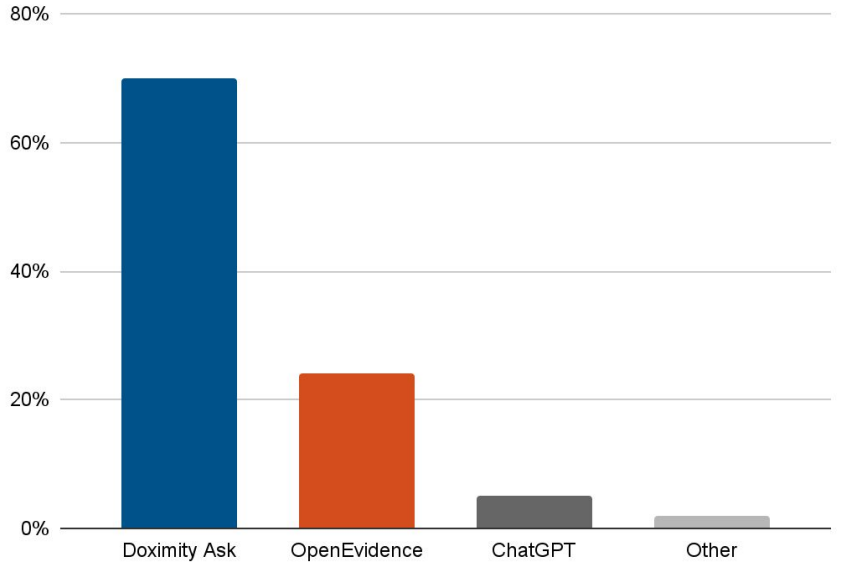
Doximity Ask was selected as the best clinical answer in 69% of side-by-side evaluations.

OpenEvidence accounted for most alternative selections (24%), followed by ChatGPT (6%)

69%

Doximity Ask was selected as the best clinical answer

Which AI answered your clinical question the best?
(By tool) | n= 4,688



Methods

A total of 4,688 independent side-by-side evaluations were completed between October 31, 2025 and April 6, 2026. During these evaluations, residents compared Doximity Ask with an alternative clinical AI tool of their choosing for clinical reference questions they posed based on their training.

All respondents were first-time Doximity Ask users, and were not incentivized to answer in any particular way.

For each comparison, respondents selected:

1. The tool providing the better clinical answer
2. The magnitude of difference ("Nearly the same", "Bit better", "Much better", "Far better")
3. A free-text explanation of their choice

Our goal is more transparency in clinical AI. If interested in specific comparative examples or broader access to this dataset, please inquire at: **pr@doximity.com**

Limitations

Results are based on self-reported user assessments and may be influenced by familiarity or interface differences. Qualitative themes reflect explicit mentions and may underrepresent implicit judgments.

Participants received compensation for completing evaluations, which may influence response behavior. In addition, all respondents were members of the Doximity platform. However, Doximity membership includes over 80% of U.S. clinicians, making this sample broadly representative of the resident population likely to encounter AI-based clinical reference tools.