

THIA User Testing Summary

User Experience Evaluation Report | JULY 2026

THIA (Trusted Health Information Assistant) is a topic- and domain-independent AI assistant currently designed for caregivers of older adults. THIA is used exclusively in research study contexts.

In July 2026, NORC conducted a structured user testing study with six family caregivers to evaluate THIA's usability, trustworthiness, and relevance ahead of broader deployment. To objectively assess usability, participants completed the System Usability Scale (SUS), a validated 10-item questionnaire that measures how easy and intuitive a system is to use across dimensions like learnability, efficiency, and user satisfaction.¹ This measure provides an objective benchmark for comparing THIA's usability against industry standards (average score: 68) and tells us whether users can effectively navigate and interact with the platform without frustration or confusion. This report summarizes what we learned and what we are doing to improve THIA.

Who Did the Testing

Six family caregivers of older adults participated in structured sessions. Participants represented a range of:

- Ages (35–65+), with 50% age 65 or older
- Racial and ethnic backgrounds (White, Black, Asian American, Hispanic)
- Caregiving experience, from under 6 months to more than 5 years
- Technology comfort levels, with 67% having prior AI tool experience

Each participant completed a pre/post survey, hands-on interaction with THIA using their own real caregiving questions, and a 30-minute follow-up interview.

What Was Evaluated

Five dimensions, each rated on a 5-point scale:

- **User Experience:** Interface usability, navigation, and comfort (4.37 out of 5)
- **Trust & Credibility:** Information accuracy, source citations, and confidence (4.21 out of 5)
- **Usefulness & Relevance:** Content quality, tone, and caregiver relevance (4.31 out of 5)
- **Technical Aspects:** Response speed, reliability, and system performance (4.72 out of 5)
- **Privacy & Security:** Data handling, comfort with sensitive information (4.46 out of 5)

What These Scores Mean

4.50–5.00 | Excellent: Exceeds expectations; a standout strength of the platform

4.00–4.49 | Strong: Performs well; meets user expectations with minor room for improvement

3.50–3.99 | Adequate: Functional but noticeably limited; targeted improvement needed

Below 3.50 | Needs Work: Meaningful gaps affecting user experience; priority improvement areas identified

Performance at a Glance



80.42 / 100 SUS score, "Excellent," exceeding the industry benchmark, top 10%



+1.33 pts gained between baseline AI trust (2.67) & trust after THIA use (4.00 out of 5)



91.7% success rate on caregiver emotional support, THIA's current focus area

What We Found

Strengths across all five dimensions. Technical performance was the highest-rated category (4.72 out of 5), with near-perfect scores for message comprehension and reliability. Privacy was rated strong (4.46 out of 5). Tone and clarity received the highest individual scores across the entire survey (both 4.83 out of 5).

THIA overcomes AI skepticism. Participants started with low baseline trust in AI-generated health information (2.67 out of 5). After interacting with THIA, trust rose to 4.00, driven by NORC's organizational backing and THIA's topical focus.

THIA excels at its core mission. The platform achieved a 91.7% success rate on caregiver emotional support and mental health resources — its strongest content areas.

Citations are trusted but invisible. Citations appeared in 92% of responses but had a 0% click-through rate; most users never noticed them. When citations were noticed, they were the strongest individual trust driver (4.50 out of 5).

What We're Doing About It

- **Improve citation visibility:** Add in-text reminders and test alternative placements so users consistently see and use source citations
- **Add proactive guidance:** Train THIA to offer alternatives when specific information is unavailable (e.g., "I don't have that, would you like these other resources instead?")
- **Redesign relevance score:** Add explanatory text and reposition to avoid overlap with the typing interface
- **Improve chat window layout:** Make chat history collapsible to widen the main window and improve readability across devices
- **Expand THIA's knowledge base:** Continuously add and refine content to increase topic coverage, improve response depth, and strengthen the platform's value for users

¹ Brooke, J. (1996). SUS: a "quick and dirty" usability. Usability evaluation in industry, 189(3), 189-194.