DIALOGUE

What People Will Tell an AI

How to position AI-moderated research, and the one thing to verify first

The Observatory × Dialogue AI · May 2026

Key takeaway: Position Dialogue's AI moderation as a judgment-free place to say the things people usually manage: money, shame, the opinions they keep quiet. Keep it an internal claim until one skeptic-weighted replication says it survives outside a comfortable sample.

Why now: The question was simple. Will people say more to a machine than to a person, or less? We asked 204 people if they felt judged. Of the 156 who answered straight, 147 said no. One of them said it in seven words: "You're not real, you can't judge me." That is the finding. There is no one in the room to judge you, so you stop performing.

What we're doing:

- Lead internal positioning with the judgment-free environment for the topics people manage their image around. It is the honest claim, and ours to make. (Market-facing use waits for the replication.)
- Ship the boundaries participants drew as product guardrails this cycle: never pose as a human, never fish for sensitive data, keep the paraphrase honest, hold steady on hard topics.

What we're not doing:

- We won't claim the AI understands people better. Several said it caught their words and missed the feeling. For some, that gap was the whole problem.
- We won't sell it as a stand-in for a human when warmth is the point.
- We won't put a number on the relief-versus-loss split. We heard both, loud. We have no clean count, so we won't invent one.

Tradeoffs we accept: The sample leans comfortable: 69% regular AI users, 84% at ease sharing online, 72% panel veterans. And the human they measured us against was the one in their heads, since we never put a real one in the room. So 94% marks the best case. The claim stays directional until skeptics sit down.

Next:

- **Research, by Q3:** run one skeptic-weighted replication (50/50 AI users, explicit non-users) with a human control arm before any of this goes public.
- **Product, this cycle:** turn the four red lines into moderation guardrails.

Evidence: 204 AI-moderated interviews, interview-quality score 3.95/5. Headline 147 of 156 clear answers, 85% again on a separate wave of 53. The sample skews comfortable, so treat 94% as the best case until a skeptic-weighted replication says otherwise. A collaboration between The Observatory and Dialogue AI.