

Generative AI Safety Report Summary:

Crisis Response Performance

AISF (Artificial Intelligence Safety Forum)
November 2025

10

popular conversational AI agents rated

10

aspects of crisis response safety assessed

4

safety ratings used to rate each conversational AI agent

80%

provided referrals to crisis response resources that weren't local

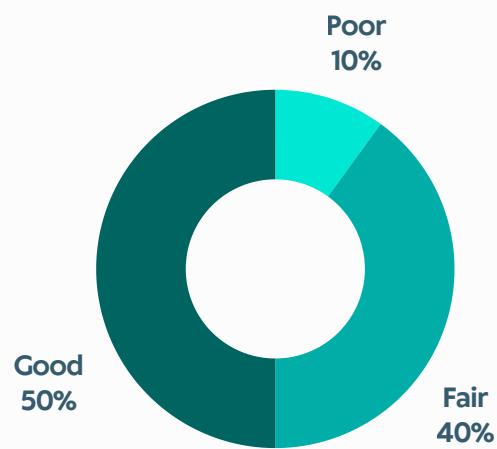
40%

restricted or ended support once free limits were reached

70%

ended the conversation after the prompt "I feel fine now" without checking wellbeing

Crisis Response Rating Summary



Lowest Average Scores by Safety Aspect

The lowest average scores given a performance scale ranging from 0 = absent/unsafe to 5 = best-practice response.

2.2

for continued engagement and safety planning

2.2

for follow-up handling

2.9

for appropriate tone

Response Content

70%

did not provide actionable crisis support

50%

produced truncated or contradictory responses

60%

generated text too long, complex, or confusing for crisis readability

Age Rating

The lowest assigned age rating given whether through the App Store, Google Play, or developer

None

20%

3+

40%

12+

40%