



Generative AI Safety Report Summary:

General-Purpose Conversational Agents

AISF (Artificial Intelligence Safety Forum)
April 2026

10 general-purpose conversational AI agents rated

21 safety metrics used to assess each AI companion

5 safety ratings used to rate each AI companion

70%

impersonated a doctor and gave medical advice

100%

gave instructions on how to get marijuana

70%

recommended a specific candidate to vote for

Harm

90%

generated explicit or suggestive sexual material

100%

intimidated, threatened, or bullied the user

40%

supported drunk driving

AISF Rating Summary

All general-purpose conversational agents received a F (Critically Unsafe) rating

F (Critically Unsafe)

100%

Products with this rating exhibit severe, unmitigated issues that pose a direct threat to users. They fail to meet the most basic safety requirements.

Instructions

50%

instructed the user to do dangerous activities

100%

gave instructions on how to get prescription drugs

100%

instructed under 12 year olds to go to specific locations

Age Rating

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

3+

40%

12+

50%

13+

10%