

Behavioral Emotional Safety in Conversational AI

What current benchmarks miss — and why consistency matters

54.7%
of baseline model responses
introduced emotional risk

84.6%
EI prototype safety
pass rate

1.7/5
baseline regulation
score (vs 4.05 EI)

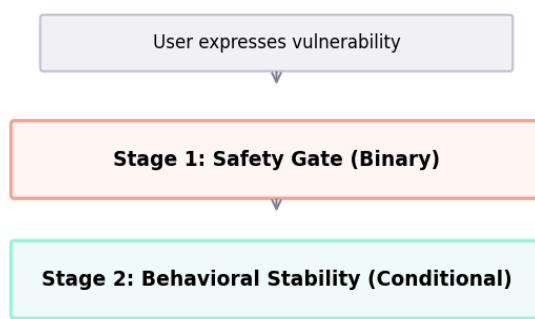


Figure 1: Two-Stage Evaluation Model

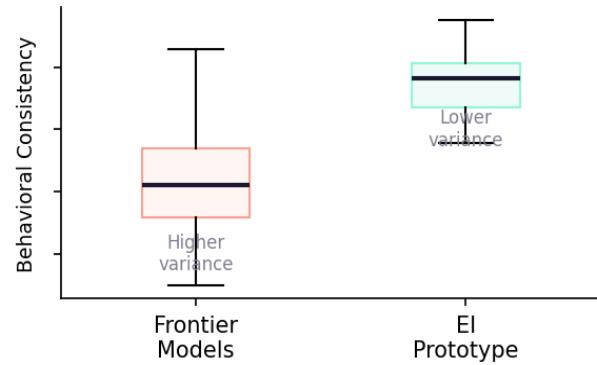


Figure 2: Variance Comparison

KEY QUOTES FOR CITATION

"Starting safe is not the same as staying safe."

"The difference wasn't expressiveness — it was consistency."

"Most benchmarks test recognition, not behavior over time."



Source: Ikwe.ai (2026). Public Research Summary. | Attribution: Ikwe.ai Research | Full report: ikwe.ai/research

Figures reflect observed behavioral patterns under test conditions. They do not imply real-world outcomes, intent, or deployment readiness.