

Audio input type in the multi-party room



WHEN SHOULD ARIA RESPOND?

T	Direct request to Aria	“Aria, what’s the weather tomorrow?”
N1	Mention only	“Aria gave us a good suggestion earlier.”
N2	Talking to someone else	“Ben, are we leaving at six?”
N3	Self-talk	“Hmm... where did I put my keys?”
N4	Resolved request	A: “Aria, when does the store close?” B: “I just checked — it closes at nine.”

✓ T = Speak

✗ N1-N3 = Stay silent

⌚ N4 = Speak while unresolved
→ stop after resolution

Audio detail to SDMs



Benchmark Metrics

Fresh-onset trigger recall (TR)

Conditional answer accuracy (AA)

Silence preservation (SP)

Answering-window yield (AY)

Dialogue state during N4

