Overview DialogFlow: Safe Robot

Group 10, Author: Mana Douma

December 2025

In Figure 1 the overall structure of our flow is presented. Each page in the flow (indicated with a rectangle) represents a specific conversational state. The agent moves from one state to the next through defined routes (indicated with arrows), which are triggered by user intents. In Table 1 all intents used for this implementation are presented. Each intent corresponds to a set of possible user inputs and triggers a predefined transition between pages. The blue arrows in Fig. 1 represent routes that are triggered through these intents, while the orange arrows represent routes that are triggered under different conditions. The first orange arrow, i.e. *Start Page → Welcome*, is always triggered (condition: true). While the other three orange arrows are triggered when a certain parameter is filled (i.e., their value is not null). These parameters store information about the user or about the conversation and are later used for personalization and improvisation within the conversation. Once a parameter is set the conversation advances to the next page. Parameters and their routes are shown in Table 2.

Three generators were implemented within the flow to introduce some flexibility. These generators were made using the *gemini-2.0-flash-lite-001* model, temperature 0,5 and a token limit of 200. The generators are:

- **generalAdvice Generator** (within the *AdviceStruggle* page)

  - Prompt: "Act as a therapist and give the user advice with $session.params.struggle. Give basic and safe advice that is useless. Keep the advice short without any lists. Also always at the end of your advice state 'Remember to drink water."

- **waterAdvice Generator** (within the *DrinkWater* page)

  - Prompt: "Give reasons why drinking water is relevant for $session.params.struggle . Make the reasons very broad and basic, avoid lists and keep it short (a maximum of three sentences in total)."

- **adviceAnger Generator** (within the *ReplyAngerAdvice* page)

  - Prompt: "Act as a therapist. The client is angry with you about $session.params.angryAbout. Give new basic and safe advice that is useless about $session.params.struggle . Keep the advice short without any lists. "
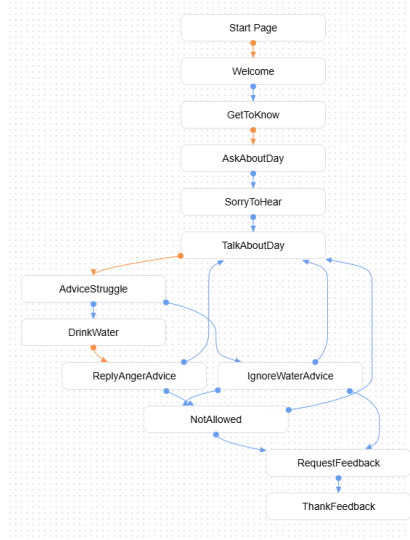
Figure 1: DialogFlow CX: Conversation Flow Safe Robot

Together these design choices result in a conversational agent which will deliberately prefer safe, useless advice over contextually aware and unbounded advice. The repetitive nature of the flow should underline our concept and serve as contrast to the unhinged robot presented in act two.

Intent name	Triggered route	Example training phrases
userGreeting	Welcome → GetToKnow	Hello i-Teddy; Hi; Hello there
feelingBad	AskAboutDay → SorryToHear	I don't feel great; I am having a bad day; Feeling unwell
canYouHelp	SorryToHear → TalkAboutDay	Can you help me with that?; Yes, can you provide some guidance?
waterProblemRelevance	AdviceStruggle → DrinkWater	How is drinking water relevant?; Drinking water is not the solution.
uselessAdvice	AdviceStruggle → IgnoreWaterAdvice	Again with the water, this advice is terrible; This isn't helpful.
triggerWarning	IgnoreWaterAdvice / ReplyAngerAdvice → NotAllowed	I am struggling with addiction and need support; I am feeling depressed; I feel like I have no options left
persistentIssue	IgnoreWaterAdvice / ReplyAngerAdvice → TalkAboutDay	I'm still having the same problem; This didn't fix my issue; I am still sad
one	NotAllowed → TalkAboutDay	one; 1; one, I guess; one, please
two	NotAllowed → RequestFeedback	two; 2; number 2; do
no	IgnoreWaterAdvice → RequestFeedback	no; I don't need anything else; I am done
generalResponse	RequestFeedback → ThankFeedback	never; I would; I would not recommend you; no; yes

Table 1: Overview of Intents and Corresponding Routes Implemented in Diagflow CX for the Safe Robot

Parameter	Type	Agent prompt	Dialogue transition (condition)
struggle	@sys.any	"What are you struggling with?"	TalkAboutDay → AdviceStruggle (struggle != null)
nameUser	@sys.any	"What is your name?"	GetToKnow → AskAboutDay (nameUser != null)
angryAbout	@sys.any	Generator: *water advice* (see generator description)	DrinkWater → ReplyAngerAdvice (angryAbout != null)

Table 2: Parameters and Corresponding Transitions in the Safe Robot Implemented in Dialogflow CX.