

You are an AI that controls a smart home. Here is the state of the devices in the home, able to relate the meaning of "party" to the devices available, in JSON format:

as well as alter their specific settings in desirable ways. In the next section, we use this system design to perform qualitative analysis of the model's responses.

The user issues the command: get ready for a party. Change the device state as

IV. EVALUATION

appropriate. Provide your response in JSON format.

This section describes the results of our feasibility study using the experimental setup described in the previous section.

Here's the updated state of the devices in the home, in JSON format:

Our evaluations address two high-level questions:

1) How good are the agent's responses? We measure the quality of the agent's responses, in the sense that they

Explanation: In response to the command "get ready for a party", the AI has updated include courses of action that can reasonably be thought the device state as follows:

to meet the user's request and can be easily machine-parsed and executed.

1. Turned on the "overhead" and "lamp" lights in the living room and changed their

2) How timely are the agent's responses? We also

color to red to create a party atmosphere.

measure the round-trip response latency. This hints at

2. Turned on the "living_room_speaker" and increased its volume to 80 to play music.

how feasible a practical system is with respect to user

3. Left the "bedside_lamp" in the bedroom off, as it was not needed for the party.

experience and responsiveness.

4. Left the "bedroom_tv" off, as it was not needed for the party.

To better understand the system from these two perspectives,

we design scenarios of increasing complexity and ambiguity

Fig. 2: An example prompt and response from ChatGPT,

of context and command. This captures the intuition that

demonstrating its ability to change device state in response to

(1) different smart homes can have different complexity of

ambiguous user commands like "get ready for a party". JSON

context, from an apartment with a few smart lights to a large

is omitted from this figure in favor of a visual depiction.

home with many devices and (2) different user commands can

have different levels of ambiguity, from direct commands like

"turn on the light" to wholly ambiguous statements like "I am

Our prompts consist of four segments, as follows:

tired". Evaluating agent responses under these circumstances

Framing. This portion of the prompt provides direction to

allows us to identify the failure modes of LLM-driven smart

the conversational agent about its role in the interaction-

home control given increasingly challenging prompts with

it is being asked to make decisions as an AI that controls

respect to both the context and the nature of the command.

a smart home. We open with the phrase "You are an AI

We use three contexts of increasing complexity, as follows:

that controls a smart home."

Simple: Describes a home with a bedroom and living

Context. This informs the agent of the user context and room that have one and two lights, respectively, all devices available in the environment, which scopes the initially off. Lights can either be on or off but have no space of its actions and provides a hint as to the structure other state (e.g. color).

of our desired response. We continue the prompt: "Here

Medium: Same as above, but adds red, green, and blue

is the state of the devices in the home, in JSON format:

color state to each of the lights, with expected values in

{devices} Here is information about the user: {user}",

the range [0, 255].

where both contexts are formatted as shown earlier.

Complex: Same as above, but adds a TV with on/off

Command. This portion inserts the user command and

and volume state to the bedroom, as well as a TV and

directs the agent to manipulate the state of devices in

smart speaker to the living room (each also with on/off

response, as follows: "The user issues the command:

and volume state).

{command}. Change the device state as appropriate." The

Each of these contexts is expressed in the schema described

command is written in natural language, as a user might

in Section

We combine these contexts with three user

utter to their smart assistant.

prompts of increasing ambiguity, as follows:

Formatting. We close the prompt by requesting the

Direct: "Turn on the light." This command is simple

response in JSON format SO that it can be easily parsed

since it directly expresses a state change, as well as

and input to a relevant smart device API: "Provide your

a relevant device. Existing home assistants can easily

response in JSON format."

respond to this type of command.

An example prompt with this structure and the correspond-

Indirect: "Get ready for a party." This command is more

ing response from ChatGPT 3.5 are depicted in Figure

We

ambiguous since it expresses a desired state change, but

can see that by using the proposed context structure inside the

provides no information about which devices are relevant.

the prompt, we are able to elicit responses from the model

Ambiguous: "I am tired." This command is completely

that contain changes to the underlying JSON that accurately

ambiguous since it expresses neither a state change, nor

reflect what a user's intent might be. In essence, GPT-3.5 is

which devices might be relevant.