Context

Command

Avg. Quality

Avg Latency (sec)

Simple

Direct

1.00

2.42

context will be necessary. This will also become essential

Indirect

0.67

2.31

as the amount of context grows to include sensor data, user

Ambiguous

0.00

2.22

preference data, and a growing and more diverse set of

Medium

Direct

1.00

4.56

controllable devices. We note that in our experiments, we did

Indirect

0.63

4.70

Ambiguous

0.17

4.97

not attempt to test the limits of how much context a model

Complex

Direct

1.00

7.90

can receive before the quality or latency of responses degrades

Indirect

1.00

7.25

substantially. This should be considered in future work.

Ambiguous

0.00

7.04

Robust system design. While we were able to leverage

Complex

Ambiguous

1.00

7.49

a simple system design in this paper, an end-to-end system

Ambiguous*

1.00

8.09

will need a more robust design to account for several factors.

TABLE I: Results for experiments given various combina-

First, since LLMs do not yet "know what they don't know",

tions of different context complexity and command ambiguity.

the likelihood of invalid or low-quality responses remains

Higher quality responses suggest the model produced a course

high. In the case of responses where the model makes invalid

of action that would be desirable for an end user (e.g., turning

changes to device state (e.g, to add new settings to a device),

on the bedroom light when receiving the command "I am tired

a full system should include a way to enforce a set of formal

and I want to sleep"). Lower latency suggests better system

properties for device states. In the case of unsatisfactory

responsiveness.

responses, it would be beneficial to develop a method for

learning user preferences or seeking clarifying information

(e.g., "are you tired and want to sleep, or are you tired but

The teaser figure depicts the result when issuing the com-

need an energy boost?").

mand "set up for a party". We include the JSON context of

From commands to automation. Our primary focus in

the light group

along with a field for the plug powering

this exploratory study was on immediate commands-the user

the stereo. The model mutates the parameters in the JSON to

makes a request and the model immediately responds with

change the stereo state to "on" and, impressively, also changes

a state change. Future work could investigate the use of the "effect" parameter of the Hue light group from "none" to

LLMs for more intuitive automation planning. A user could,

"colorloop" to create a looping color effect. The latter change

for instance, ask their smart assistant to "play jazz when it

suggests that GPT-3 may have been trained on material about

rains" and the model could leverage contextual information to

the specific features of the Hue API and can leverage that

put in place an automation sequence that meets their needs.

along with the inferred meaning of the user command to

This would obviate the need for pre-programmed automation

trigger more intuitive changes than existing systems.

routines and could substantially improve user satisfaction with

We briefly list multiple other commands we tested in our

smart assistant systems.

implementation, along with responses from the model:

VII. CONCLUSION

"make it bright in here" - sets lights to full brightness

"make it groovy" - sets lights to color loop; adds invalid

In this paper, we explored the feasibility of smarter smart

"genre" field to stereo and sets it to "groovy"

home control using large language models (LLMs). We pro-

"gotta relax" - dims lights, turns stereo on

posed a simple system design for capturing smart home con-

"I'm cold" - sets lights to warm white, turns stereo on

text (i.e., information about the user and controllable devices

"I'm leaving" - turns off lights and stereo in the environment) in engineered prompts to GPT-3, showing

"I'm home" - turns on lights and stereo

that the model has the ability to infer meaning behind indirect

and ambiguous user commands like "I am tired and I need to

We note, of course, that these tests are far from exhaustive.

work" and, in response, generate changes to smart device state.

We observed high variability in responses, meaning the same

We implemented our system design, giving GPT-3 control

command can elicit many responses: some good, some bad.

of real devices and finding that it is able to quickly and

A more robust system design will be necessary to tackle the

appropriately control them in response to user commands with

inconsistencies present in current LLM model outputs. We

no fine tuning and no post-processing of its responses. By

address this in our discussion in the following section.

simply telling GPT-3 what devices are available and what the

VI. DISCUSSION & FUTURE WORK

user wants, it can generate courses of action in response.

Our work hints at the capability of GPT-3 and similar

Our efforts in this paper hint at exciting opportunities for

models to go far beyond the current abilities of smart space

future work. We suggest several avenues for further research.

Managing contextual information. We found that includ-

control and motivates future work with context modeling, end-

to-end system design, and approaches for further leveraging

ing more context can improve the quality of the model's

GPT-3's capabilites to develop complex automation routines

responses, but at the expense of response latency. To effec-

in response to user commands.

tively navigate this tradeoff in an end-to-end solution, a more

involved approach for storing, pre-processing, and expressing

https://developers.meethue.com/develop/hue-api/groupds-api/