domain contextual reasoning be applied to practical smart
tasks, such as text classification and sentiment analysis. In
home applications?
the same year, OpenAI proposed GPT (Generative Pre-trained
To explore this question, we carry out a feasibility study that
Transformer) 22|. Both models use a transformer architec-
places GPT-3 in control of a smart home. We evaluate GPT-
ture 27 that was pre-trained on a massive corpus of text data,
3's ability to provide high-quality responses to user commands
including books, articles, and websites. The resulting models
of varying ambiguity given only a simple prompt and a data
demonstrate impressive results on a wide range of natural
structure containing information about devices that it can con-
language processing tasks, including language translation,
trol. Our results demonstrate that LLMs like GPT can infer the
text generation, and the ability to translate natural language
meaning behind ambiguous user commands like "get ready for
descriptions into program implementations.
a party" or "I am tired and I want to sleep" and respond with
Following the success of the transformer-based model, sub-
properly-formatted data describing courses of action, enabling
sequent studies have explored ways to improve and expand
more intuitive control of smart devices. We furthermore build
the model's performance. In 2019, Radford et al., published
a proof-of-concept implementation that puts GPT-3 in control
an updated version of GPT and called GPT-2 [23]. Building

of real devices, showing LLM-driven command inference and

on the success of GPT-2, Brown et al. released GPT-3 in 2020

action planning can function in practice with no fine-tuning or

4 After that, in 2023, GPT-4 was introduced. It is currently

task-specific training required. Motivated by our results, we

one of the largest and most powerful language models, with

propose future work that can further leverage the power of

more than 1 trillion parameters [18]. At time of writing, access

LLMs toward building smarter smart home applications.

to GPT-4 is limited-we therefore base our study on GPT-3.

Our key contributions are as follows:

Two popular approaches exist for adapting task-agnostic

An experimental setup and study results that show LLMs

LLMs to new applications: prompt engineering and fine-

can infer meaning behind abstract user commands like "I

tuning. Prompt engineering refers to the process of designing

am tired and I have to work" and, in response, quickly

a task-specific prompt or template that guides the model to

and appropriately change the state of the smart devices

produce relevant outputs for a particular task 29]. These

available in the home, with no task-specific training.

prompts generally contain instructions to the model written

An implementation that puts a GPT model in control of

in natural language-e.g., "explain the following passage

real devices, showing that it can intuitively respond to a

of text". Fine-tuning, on the other hand, involves directly variety of commands. When told to "set up for a party", it

training the model on a new task by providing task-specific

responds by turning on a stereo and configuring a group

examples [23]. The key advantage of prompt engineering over

of Hue lights to loop through a festive set of colors;

fine-tuning is that it does not require task-specific data-

given the command "I'm leaving", it turns off all available

we therefore adopt that approach here. Within the realm of

devices. We trigger these actions by inputting the LLM's

prompt engineering, there are zero-shot and few-shot learning

response directly into smart device APIs.

approaches. Zero-shot approaches provide the model with a

Analytical results that suggest responses are variable in

single prompt containing instructions and task-specific infor-

quality, dependent on both the devices available and the

mation; few-shot approaches provide examples to the model

nature of the user's command. In essence, further system

of correct input/output pairs. We focus on zero-shot learning.

design is necessary to manage the LLM's tendency to

Context-aware spaces leverage sensor information, user

"not know what it doesn't know" in order to produce

data (including past behaviors and preferences), and de-

consistently high-quality responses.

vice state to influence system actions toward meeting user

The following describes the structure of this paper. Sec-

needs 2 The notion of "context-awareness" in this sense tion

situates our work with related research. Section

has roots in research on ambient intelligence I-that is, the

describes the experimental setup that we use to demonstrate

development of built environments that sense and adapt to

the feasibility of LLMs as smart home controllers. Section IV

users. A concrete example of this concept is a home that lever-

presents the results of our exploratory study, while Section

ages contextual information to improve energy efficiency 11

demonstrates a proof-of-concept implementation. Section

7 In an early paper, Yamazaki suggested that smart homes

offers avenues for future work. Section VII concludes.

should go beyond automation and instead integrate expressive

interfaces between the user and system [28], a goal that is

II. BACKGROUND & RELATED WORK

partially realized in smart assistants [20], but with limited

This section provides a high-level overview of LLMs and

ability to adapt to more complex user commands [13] Ample

their applications before situating our work with related efforts

prior work has approached the issue of context-awareness us-

in context-aware smart spaces.

ing task-specific models 21 12 17 14| While these methods

Large language models (LLMs) have gained significant

can achieve high performance given ample task-specific data,

attention in recent years due to their impressive performance

we believe that the high zero-shot performance of LLMs could on a wide range of natural language processing tasks. In

hint at better generalizability without a need for training data.

2018, Devlin proposed BERT, a language representation model

However, we are aware of no work to-date that has explored

that uses Bidirectional Encoder Representations from Trans-

the use of task-agnostic LLMs for deeper contextual reasoning

formers

6

and can be fine-tuned for a variety of NLP

in smart environments. This motivates our feasibility study.