?Get ready for a party?: Exploring smarter smart spaces with help from large language

models Evan King, Haoxiang Yu, Sangsu Lee, Christine Julien The University of Texas at

Austin {e.king, hxyu, sethlee, c.julien}@utexas.edu 3202 raM 42]CH.sc[GPT models can

infer meaning behind ambiguous user commands and control smart home devices in

response. When told

to ?setupforaparty?,GPT-3producesvalidJSONthatsetsthelightstoacolorloopandturnsonthestereoformusicplay

right response to someone who says ?get ready I. INTRODUCTION for a party? is deeply

influenced by meaning and context. For a smart home assistant (e.g., Google Home), the

ideal response An exciting prospect of smart homes at their advent

was mightbetosurveytheavailabledevicesinthehomeandchange the potential to reduce user

burden by providing seamless, their state to create a festive atmosphere. Current

practical unobtrusive,and?smart?interfacestoeverydaydevices.While systems cannot service

such requests since they require the ability to (1) infer meaning behind an abstract statement

and smart assistants have improved significantly over the years (2)mapthatinferencetoaconcretecourseofactionappropriate with respect to speech

recognition [25, 24] and user satis- for the context (e.g., changing the settings of specific

devices). faction [20, 16], a central challenge remains: how can these In this paper, we

leverage the observation that recent

task- assistantsbemadetorespondappropriatelytoambiguoususer agnostic large language

models (LLMs) like GPT-3 embody

a commandsthatmaybeinfluencedbycontextorareotherwise vastamountofcross-domain,sometimesunpredictab

that existing rule-based home assistant systems lack,

impossibleforsystemdeveloperstoanticipatebeforehand?An which can make them powerful

tools for inferring user intent example of such a command might be a user preparing their and

generating appropriate context-dependent responses during home to entertain for guests, who

asks their smart assistant to smart home interactions. We first explore the feasibility of a ?get

ready for a party?. The hope is that the assistant?if it is system that places an LLM at the center of command inference truly smart?might be able to help by inferring the meaning and action

planning, showing that LLMs have the capacity

to inferintentbehindvague,context-dependentcommandslike?get of the statement and

determining how to change the state

of readyforaparty?andrespondwithconcrete,machine-parseable available devices in response:

perhaps to start up the user?s instructions that can be used to control smart devices. We

partyplaylistonasmartspeakerandchangetheirsmartlights furthermoredemonstrateaproof-of-conceptimplemen

a festive color scheme. In practice, however, such a

request putsanLLMincontrolofrealdevices,showingitsabilitytoinfer isbeyondthecapacityofcurrentsmarthomes

and change device state appropriately with no

fine-tuning ortask-specifictraining.OurworkhintsatthepromiseofLLM- Home will sadly

admit: ?I?m sorry, I didn?t understand.? driven systems for context-awareness in smart

environments, Inthispaper,wearemotivatedbytheobservationthatlarge motivating future research in this

area. languagemodels(LLMs)likeOpenAI?sGPT-3[3]haveshown animpressiveabilitytogeneralizetonewtasksv

performance, as well as the capacity to infer meaning Index Terms?human-centered

computing, artificial intelli- behind semantically complex or abstract statements [15].

We gence, internet of things thus ask the question: can this powerful capacity for

cross-domain contextual reasoning be applied to practical smart tasks, such as text

classification and sentiment analysis. In home applications?

thesameyear,OpenAIproposedGPT(GenerativePre-trained Toexplorethisquestion,wecarryoutafeasibilitystudy

Transformer) [22]. Both models use a transformer architec- places GPT-3 in control of a

smart home. We evaluate GPT-

ture[27]thatwaspre-trainedonamassivecorpusoftextdata, 3?sabilitytoprovidehigh-qualityresponsestousercomm

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 structurecontaininginformationaboutdevicesthatitcancon- language processing tasks,

including language translation, trol.OurresultsdemonstratethatLLMslikeGPTcaninferthe text

generation, and the ability to translate natural

language meaningbehindambiguoususercommandslike?getreadyfor 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-formatteddatadescribingcoursesofaction,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

onthesuccessofGPT-2,Brownetal.releasedGPT-3in2020 actionplanningcanfunctioninpracticewithnofine-tuni

[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 morethan1trillionparameters[18].Attimeofwriting,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 ? Anexperimentalsetupandstudyresultsthatshowl LLMs LLMs to new applications: prompt engineering and fine- caninfermeaningbehindabstractusercommandslike?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 varietyofcommands.Whentoldto?setupforaparty?,it training the model on a new task by

providing task-specific responds by turning on a stereo and configuring a group

examples[23].Thekeyadvantageofpromptengineeringover of Hue lights to loop through a

festive set of colors; fine-tuning is that it does not require task-specific

data? giventhecommand?I?mleaving?,itturnsoffallavailable 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 needs [2]. The notion of ?context-awareness? in this sense The following describes the structure of this paper. Sec- has roots in research on ambient intelligence [5]?that is, the tion II situates our work with related research. Section III development of built environments that sense and adapt to describes the experimental setup that we use to demonstrate users.Aconcreteexampleofthisconceptisahomethatlever- the feasibility of LLMs as smart home controllers. Section IV ages contextual information to improve energy efficiency [11, presents the results of our exploratory study, while Section V 7]. In an early paper, Yamazaki suggested that smart homes demonstrates a proof-of-concept implementation. Section VI should go beyond automation and instead integrate expressive offers avenues for future work. Section VII concludes. interfaces between the user and system [28], a goal that is II. BACKGROUND&RELATEDWORK 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 theirapplicationsbeforesituatingourworkwithrelatedefforts prior work has approached

the issue of context-awareness us- in context-aware smart spaces.

ingtask-specificmodels[21,12,17,14].Whilethesemethods 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

webelievethatthehighzero-shotperformanceofLLMscould on a wide range of natural language

processing tasks. In hint at better generalizability without a need for training

data. 2018,DevlinproposedBERT,alanguagerepresentationmodel However, we are aware of

no work to-date that has explored that uses Bidirectional Encoder Representations from

Trans- theuseoftask-agnosticLLMsfordeepercontextualreasoning formers [6] and can be fine-tuned for a variety of NLP in smart environments. This motivates our feasibility study.III. SYSTEMDESIGN In this section, we introduce the system design that we use to explore the feasibility of LLM-driven smart home control. We first assume the use of an LLM like GPT-3 that provides responses to user prompts written in natural language. These LLM models are not task-specific, rather, they are trained on an immense amount of cross-domain textual information and, depending on the structure of the prompt, can provide Fig. 1: Data structures for expressing smart home device and responsessuitedtoavarietyofdifferentusecases(e.g.,writing user context in prompts to an LLM. a poem, writing code in response to a high-level program description, etc.). We opt to adapt the model?s outputs to our task using zero-shot learning through prompt engineering. color values. This overall structure is depicted in Fig. 1 and Our challenge is therefore to package relevant context and illustrated by the example in the following: user commands into a concise prompt issued to the model, { such that its responses include concrete, machine-parseable "user":

{ changestodevicestatethatcanbepassedofftotheappropriate "location": "living_room" smart

device APIs. Qualitatively, we want these courses

of } actiontobeshapedbythemodelsuccessfullyinferring(1)the } intent behind the user

command and (2) the manner in

which thestateofavailabledevicescanbechangedtomeettheuser?s { intent. To that end, we first

define an abstract schema for "devices": { capturing smart home context before describing a

method for "bedroom": { engineeringpromptstoconversationalLLMsthatelicituseful, "lights":

{ actionable responses. "bedside_lamp": { "state": "off" A. Context Schema

} } Inorderforthemodelto?know?whatactionsareavailable }, toit,weneedtopackagetheavailabledevices,theirsta

{ other relevant information?i.e., the context?into a machine- "lights":

{ parseableformat.Thispackageeffectivelydescribestheaction "overhead": { space available to

the model: the knobs it can turn, and "state": "on" information (e.g., which room the user is in) that might }, influence how it turns them. It also provides a hint about "lamp": { how the model should format its response. Representations "state": "off" of context can be complex and have been explored in the } literature [9, 1]. Since our goal is to conduct an exploratory }, study rather than design an end-to-end solution, we use a "tvs": { schemathatissimplebutadequateforourexperimentalsetup. "living_room_tv": { We choose JSON for structuring this data since it is the de- "state": "off", factodatainterchangeformatforRESTfulAPIsusedbymany "volume": 20 smart home devices [19, 8, 10]. Leveraging a common format } is also advantageous since there is a high likelihood that the } LLM has been trained on source material that contains it, } which benefits the model?s ability to converse in

it. } Atthetoplevel,contextisacollectionof?key,value?pairs. } There are two relevant contexts: ?user? context that contains immutableinformationabouttheuser?sstate,e.g.,whichroom

Inthisexample,theuser?shomehastworooms?abedroom theyarein,and?device?context,whichcontainsmutablea andlivingroom?andtheuseriscurrentlylocatedintheliving immutable information about the

devices in the home. Each room. The bedroom has one light turned off, and the

living top-levelkeyinthecollectionofdevicecontextdefinesaroom room has two lights (one

turned on) and a television. in the home, and within each room we define collections of B.

Prompt Engineering devices organized by type, e.g., ?lights?, ?tvs?, etc. Within a collection of

devices, we define each individual device as Having developed a structure for storing

context, we now a collection of properties about that device, e.g., for a light move to the

practical challenge of engineering prompts that we can define its ?state? property and its ?r?, ?g?,

and ?b? elicit useful responses from the model.able to relate the meaning of ?party? to the devices

available, 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. IV.

EVALUATION This section describes the results of our feasibility

study usingtheexperimentalsetupdescribedinthepreviousection. 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 includecoursesofactionthatcanreasonablybethought to meet the user?s request and can be easily machine- parsed and executed. 2) How timely are the agent?s responses? We also measure the round-trip response latency. This hints at how feasible a practical system is with respect to user experience and responsiveness. Tobetterunderstandthesystemfromthesetwoperspectives, 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 ambiguoususercommandslike?getreadyforaparty?.JSON context, from an apartment with a few smart lights to a large is omitted from this figure in favor of a visual depiction. homewithmanydevicesand(2)differentusercommandscan have different levels of ambiguity, from direct commands like ?turnonthelight?towhollyambiguousstatementslike?Iam Our prompts consist of four segments, as follows: tired?. Evaluating agent responses under these circumstances ? Framing.Thisportionofthepromptprovidesdirectionto allows us to identify the failure modes of LLM-driven smart theconversationalagentaboutitsroleintheinteraction? home control given increasingly challenging prompts with itisbeingaskedtomakedecisionsasanAIthatcontrols 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 spaceofitsactionsandprovidesahintastothestructure 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}.Changethedevicestateasappropriate.?The

Eachofthesecontextsisexpressedintheschemadescribed command is written in natural

language, as a user might in Section III. 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:?Getreadyforaparty.?Thiscommandismore ing response from ChatGPT 3.5 are depicted

in Figure 2. We ambiguous since it expresses a desired state change,

but canseethatbyusingtheproposedcontextstructureinsidethe

providesnoinformationaboutwhichdevicesarerelevant. 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.We run our tests with each possible combination of these and a subjective

command like ?get ready for a party?, the three contexts and commands (9 total), each for 10

trials. We LLM simply makes up a response?specifically, it turns on save the agent?s response

for each trial in a human-readable all the lights in the house, to include the bedroom.

When format, then perform manual rating to measure the quality of we add a speaker and a television to the context, the model the responses. Our process for rating the quality of responses now has more relevant knobs to turn, and produces a higher is based on a binary label, where each is assigned one of the quality response. The tendency to make something up when following labels based on its quality: the answer is unknown or requires more context is an open problemandmotivatesthedevelopmentofapplication-specific ? Poor(0):?Thechangestothedevicesdonotatallreflect methods to mediate between the user and the model. the intent behind the user command, or the response is For the most ambiguous command (?I am tired?), we note malformed/garbled.? that the model delivers poor responses regardless of context. ? Good(1):?Thechangestothedevicesarereasonablefor In all but a few cases, the LLM simply turns on all of the the command. You can imagine someone being satisfied home?slights.TheexceptionisinaMedium/Ambiguoustrial, with the result, even if it is somewhat subjective (e.g., where it only turns on the bedroom lamp, perhaps to help the based on different personal preferences).? user prepare for bed. This is to be expected: an individual?s Threeresearchersindependentlyreviewedallresponsesand intent and preference in this case are highly subjective (are assigned them a label. We report the aggregate score for each they, e.g., tired and ready for bed or tired but they have

a trialastheaverageacrossallassignedscores.Wealsonotethe pressing deadline?) and the LLM ultimately cannot read the average latency for each trial?this includes both the network user?s mind. However, since the LLM does not ?know what it transmission time of the request, as well as the inference time doesn?t know?, it does not ask for clarification or go with the taken by the model. Since this time is subject to network safest choice, which is likely to do nothing. Instead, it makes conditions and API demand, it should be taken as a

rough something up. estimate rather than a concrete benchmark. Our results are Since our previous results suggested more context is often summarized in Table

I. beneficial,wedigdeepertoseeifwecanhelpitmakeabetter Response time is a function of context complexity. With choice. We amend the vague command ?I am tired? to offer respect to latency, we can see that responses generally arrive hints at the user?s intent: on the order of seconds, meaning that a practical system ? Ambiguous*: ?I am tired and I need to work.? could feasibly leverage an LLM for ambiguous command ? Ambiguous**: ?I am tired and I want to sleep.? inference and action planning without significant detriments Provided this added hint at the user?s context, the response to user experience or responsiveness. For direct commands, a quality improves significantly. For Ambiguous*, the model 2 to 3 second response time may be too long?future system consistently responds by turning on the living room lights designs could thus leverage the LLM only for commands that while leaving all other devices off; for Ambiguous**, the require it. This may entail a hybrid of rule-based inference model turns on only the user?s bedside lamp and, in some for common commands, along with LLM inference for less cases, reduces the volume on their speaker and TV. Note that familiar commands. It is also worth noting that as the context although the amended statement includes additional context, increases in complexity, the response latency also increases? it still requires the model to infer meaning in a way that more this motivates future work to develop methods for filtering rigid or rule-based approaches cannot. context prior to prompting the agent so that only the most relevant information is provided. V. IMPLEMENTATION Response quality is a function of context and command ambiguity. With respect to response quality, we find that the To demonstrate LLM-driven smart home control in prac- LLMapproachprovidesgoodresponsesgiventhesamedirect tice, we built a proof-of-concept implementation in Python. and simple commands that current home assistants are able to Our implementation accepts user commands as strings, pack- service. Note, however, that unlike existing home assistants, ages them into prompts along with contextual information theLLMapproachutilizesamuchsimplersystemarchitecture about real devices, then processes responses from OpenAI?s that performs command inference and action planning in text-davinci-003 model into smart home API calls that the same pass.

These results are consistent given increasing change the device state as specified by the

LLM. We scope degrees of context complexity, suggesting that the model was the

application to one room (a researcher?s living room) with not overwhelmed by the growth in

the decision space that three Philips Hue color smart lights [19] and one

TP-Link comeswithaddingnewdevicesandpossiblestatechanges.On Kasa smart plug [26] that

controls a stereo. We store the the contrary, the model provides better responses when given

device context in JSON as in our experimental setup, with more context that might be

relevant to the user?s command, the difference that the device state for the Hue lights is

pulled as is apparent when comparing the low response quality of directly from the Hue API

without modification. Our code is the Simple/Indirect and Medium/Indirect experiments

against open source and available online.1 the Complex/Indirect experiment. Upon inspection

of the responses, the reason for this is clear: given minimal context

1https://github.com/UT-MPC/homegptContext Command Avg.Quality

AvgLatency(sec) context will be necessary. This will also become essential Simple Direct

1.00 2.42 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 not attempt to test the limits of how much context a model Ambiguous 0.17

4.97 canreceivebeforethequalityorlatencyofresponsesdegrades Complex Direct 1.00

7.90 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?, tionsofdifferentcontextcomplexityandcommandambiguity. the likelihood of invalid or

low-quality responses remains Higherqualityresponsessuggestthemodelproducedacourse high. In the case of responses where the model makes

invalid ofactionthatwouldbedesirableforanenduser(e.g.,turning changes to device state (e.g, to

add new settings to a device), onthebedroomlightwhenreceivingthecommand?Iamtired 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 group2, along with a field for the plug powering

thisexploratorystudywasonimmediatecommands?theuser the stereo. The model mutates the

parameters in the JSON to makes a request and the model immediately responds

with changethestereostateto?on?and,impressively,alsochanges 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.

routinesandcouldsubstantiallyimproveusersatisfactionwith 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 In this paper, we explored the feasibility of smarter smart ? ?make it groovy? ? sets lights to color loop; adds invalid home control using large language models (LLMs). We pro- ?genre? field to stereo and sets it to ?groovy? posed a simple system design for capturing smart home con- ?

?gotta relax? ? dims lights, turns stereo on text (i.e., information about the user and controllable

devices ? ?I?m cold? ? sets lights to warm white, turns stereo

on intheenvironment)inengineeredpromptstoGPT-3,showing ? ?I?m leaving? ? turns off lights and

stereo thatthemodelhastheabilitytoinfermeaningbehindindirect ? ?I?m home? ? turns on lights and

stereo 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,inresponse,generatechangestosmartdevicestate. 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 appropriatelycontroltheminresponsetousercommandswith 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&FUTUREWORK 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. controlandmotivatesfutureworkwithcontextmodeling,end- Managing contextual

information. We found that includ- 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 2https://developers.meethue.com/develop/hue-api/groupds-api/REFERENCES

[25] Prakhar Swarup et al. ?Improving ASR confidence scores for

Alexa usingacousticandhypothesisembeddings?.In:(2019). [1] Hamim Md Adal et al. ?The

Space Broker: A Middleware for Mediating Interactions in Smart IoT Spaces?. In:

Proceedings of the [26] TP-Link Kasa Smart Plugs. 2023. URL:

https://www.kasasmart.com/ 8th ACM International Conference on Systems for

Energy-Efficient us/products/smart-plugs. Buildings, Cities, and Transportation. BuildSys ?21.

Coimbra, Portu- [27] Ashish Vaswani et al. ?Attention is all you need?. In: Advances

in neuralinformationprocessingsystems30(2017). gal:AssociationforComputingMachinery,2021,pp.101?110.

TatsuyaYamazaki.?BeyondtheSmartHome?.In:2006International 9781450391146. DOI:

10.1145/3486611.3486664. URL:

https://doi. ConferenceonHybridInformationTechnology.Vol.2.2006,pp.350? org/10.1145/3486611.3486664

MuhammadRaisulAlam,MamunBinIbneReaz,andMohdAlauddin [29] Tony Z. Zhao et al.

Calibrate Before Use: Improving Few-Shot Per- Mohd Ali. ?A Review of Smart Homes?Past,

Present, and Future?. formanceofLanguageModels.2021.arXiv:2102.09690[cs.CL]. In: IEEE

Transactions on Systems, Man, and Cybernetics, Part

C (ApplicationsandReviews)42.6(2012),pp.1190?1203.DOI:10.1109/ TSMCC.2012.2189204. [3]

Tom Brown et al. ?Language models are few-shot learners?. In:

Ad- vancesinneuralinformationprocessingsystems33(2020),pp.1877? 1901. [4]

TomB.Brownetal.?LanguageModelsareFew-ShotLearners?.In: CoRRabs/2005.14165(2020).arXiv:2005.141

URL:https://arxiv. org/abs/2005.14165. [5] Diane J Cook, Juan C Augusto, and Vikramaditya

R Jakkula. ?Am- bient intelligence: Technologies, applications, and opportunities?.

In: Pervasiveandmobilecomputing5.4(2009),pp.277?298. [6]

JacobDevlinetal.?Bert:Pre-trainingofdeepbidirectionaltransform- ersforlanguageunderstanding?.In:arXivprep

PRGeraldoFilhoetal.?Energy-efficientsmarthomesystems:Infras- tructureanddecision-makingprocess?.In:Inte

Google Nest API Documentation. 2022. URL:

https://developers. google.com/nest/device-access. [9] Jie Hua et al. ?CoPI: Enabling

Probabilistic Conflict Prediction in Smart Space Through Context-awareness?. In: 2022

IEEE/ACM Seventh International Conference on Internet-of-Things Design

and Implementation(IoTDI).IEEE.2022,pp.30?42. [10] Insteon REST API Documentation.

2022. URL: https://www.insteon. com/developer. [11] Marco Jahn et al. ?The energy aware

smart home?. In: 2010

5th internationalconferenceonfutureinformationtechnology.IEEE.2010, pp.1?8. [12] M

Humayun Kabir et al. ?Machine learning based adaptive

context- awaresystemforsmarthomeenvironment?.In:InternationalJournal ofSmartHome9.11(2015),pp.55?62

Julia Kiseleva et al. ?Understanding user satisfaction with

intelligent assistants?.In:Proceedingsofthe2016ACMonConferenceonHuman InformationInteractionandRetri

TiankaiLiangetal.?Anunsuperviseduserbehaviorpredictionalgo- rithmbasedonmachinelearningandneuralnetw

Xiao Liu et al. ?GPT understands, too?. In: arXiv preprint arXiv:2103.10385(2021). [16]

IreneLopatovskaetal.?Talktome:Exploringuserinteractionswith the Amazon Alexa?. In: Journal

of Librarianship and Information Science51.4(2019),pp.984?997. [17] Isaac Machorro-Cano et al. ?HEMS-IoT: A big data and machine learning-based smart home system for energy saving?.

In: Energies 13.5(2020),p.1097. [18]

OpenAI.GPT-4.URL:https://openai.com/product/gpt-4. [19] Philips Hue API Documentation.

2022. URL: https://developers. meethue.com/. [20]

AmandaPuringtonetal.??AlexaismynewBFF?SocialRoles,User Satisfaction,andPersonificationoftheAmazon

the 2017 CHI conference extended abstracts on human factors

in computingsystems.2017,pp.2853?2859. [21] Basheer Qolomany et al. ?Leveraging machine

learning and big data for smart buildings: A comprehensive survey?. In: IEEE Access

7 (2019),pp.90316?90356. [22]

AlecRadfordetal.?Improvinglanguageunderstandingbygenerative pre-training?.In:(2018). [23]

Alec Radford et al. ?Language models are unsupervised

multitask learners?.In:OpenAIblog1.8(2019),p.9. [24] Anirudh Raju et al. ?Scalable multi

corpora neural language models forasr?.In:arXivpreprintarXiv:1907.01677 (2019).