Embedding spatiotemporal context for conversations with autonomous mobile robots

Elliott Hauser
The University of Texas at Austin
Austin, Texas, United States of America
elliott@utexas.edu

Justin Hart
The University of Texas at Austin
Austin, Texas, United States of America
elliott@utexas.edu

Samuel Baker
The University of Texas at Austin
Austin, Texas, United States of America
sebaker@austin.utexas.edu

Luis Sentis
The University of Texas at Austin
Austin, Texas, United States of America
lsentis@utexas.edu

ABSTRACT

We present our designs for utilizing large language models for contextualized conversational interaction in the context of long term deployment of autonomous robots. The example of a university campus tour is used.

KEYWORDS

Human-Robot Interaction, Large Language Models, Autonomous robots

ACM Reference Format:
Elliott Hauser, Justin Hart, Samuel Baker, and Luis Sentis. 2023. Embedding spatiotemporal context for conversations with autonomous mobile robots. In *Proceedings of Workshop on Human-Robot Conversational Interaction (HRCI'23).* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Large language models (LLMs), deep learning algorithms trained on vast corpora, have attracted tremendous recent interest. The practical applications of LLMs range from zero-shot image recognition, as in the case of CLIP [3], to conversational interfaces, as in the case of the case of ChatGPT. Because LLMs produce grammatically and syntactically well-formed output in a range of styles and genres and are highly responsive to user prompts, HRI research looms as an additional potential application area. This prospect of using LLMs in HRI research raises major questions about best practices, potential benefits, and potential risks [2]. To begin the work of sorting out these questions, this paper elaborates on issues we anticipate may arise with the use of LLMs to enable contextual conversational interfaces to a long-term deployment of autonomous mobile robots.

Our Living and Working with Robots program at The University of Texas at Austin is focused on developing long-term deployments of service robots on the UT campus. Faculty, students, staff, and visitors to the UT campus will be able to use robots for tasks such as delivery, interacting with the library system, and taking campus tours. The robots we deploy should benefit the campus community, should be generally welcomed by it, should be responsive to the community's expressed wants and needs, and should not have incidental negative impacts on humans whom they encounter who are not otherwise involved in their tasks. To facilitate living and working with robots, we will need to make communication between humans and deployed long-term autonomous general purpose service robots more effective. The emergence of LLMs presents an opportunity to significantly upgrade such communication by providing contextual conversational interfaces between the humans and robots concerned.

Challenges to implementing such interfaces include a variety of relatively well known issues pertaining to LLMs, but moreover a set of issues we expect to emerge when LLMs are built into general purpose service robot systems that can be encountered by the public. As with LLMs generally, when we deploy an LLMin this way we will need to assure that the model generates relevant and factually-accurate statements and that we mitigate the possibility that it might generate offensive, stereotyped, or otherwise harmful output. While we will be able to draw on LLM-centered research to help us address these issues, we also hope our experience deploying LLMs in context may shed new light on the nature of these issues and thereby make contributions to LLM research. Meanwhile, embedding LLMs in contextual conversational interfaces can take HRI research in new directions.

This paper will highlight three such LLM-driven new directions in HRI research, each of which raises further challenges. First, we believe that if we leverage LLMs to build contextual conversational interfaces onto general purpose service robots, we can better integrate appropriate gesturing and other body language skills into those systems. Secondly–and this is our particular focus in what follows–we believe LLM-enabled contextual conversational interfaces may potentially mitigate the danger that such robots will negatively impact incidentally co-present persons. Finally and more broadly, integrating LLMs into HRI research might help lay the groundwork for the more effective use of large HRI-tailored datasets by facilitating the collection and analysis of HRI data at a new scale and across hitherto unconnected contexts.

HRCI'23, March 2023, Stockholm, Sweden

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

2 MOTIVATION

The authors are collaborators on an interdisciplinary team that is deploying robots to a university campus, and conceiving of this deployment as an integrated technological and social scientific research intrument. The team is especially focused upon human-robot encounters (HREs), an emerging area of study in HRI centered upon incidental encounters with robots by non-operators.

2.1 Human-Robot Encounters

A core framing of this project is that potential negative impacts of the long term use of autonomous systems will be experienced by those involved in what we call *collateral encounters*, instances of incidental copresence with robots by humans other than those operating the robots or directly benefitting from their operation. The population of such humans has been termed incidentally copresent persons (InCoPs)by Rosenthal-von der Pütten et al. [4]. An major arm of our research thus seeks to foster positive incidental encounters by enabling rich, dynamic, and contextual conversational human-robot interactions geared neither toward the use of those robots nor toward the fulfillment of the robots' purposes.

2.2 Conversational Capability and Robot Encounters

The intuition behind this approach is that the perception of the potential for rich engagement is part of the experiential character of incidentally encountering other humans in a community. Bus drivers or mail carriers, for example, are capable of engaging in polite conversation, sharing basic contextual information about an area, and giving simple directions, even (or perhaps especially) when not directly serving an interlocutor in their official capacity. The potential for such interactions with service professionals is part of what makes daily encounters with them feel routine at worst, and often comfortingly familiar. The overarching hypothesis we seek to test is that robots which can provide such potential and actual contextual conversational interaction will make a similarly neutral-to-mildly-positive impact through the many collateral encounters they will precipitate during their operation. If this hypothesis is confirmed by initial experiments, that will suggest that the construction, optimization, adaptation, and maintenance of location-specific and context-aware robots could become a central area of research in the emerging field of HREs research.

3 LARGE LANGUAGE MODELS

The term "LLMs " is not precisely defined, but usually indicates those trained on Web-scale datasets. A useful overview and critique are provided in [1].

Models are trained and made available by commercial or open source services, using large amounts of computational resources to do so. In general, models that utilize more input data and/or more generations of training are larger and more capable, while smaller versions are more performant.

Considering that our research goals (and ethical responsibilities to participants, given LLMs' potential to generate hate speech) demand customizing models' behavior, we plan to explore the relative strengths and overall utility of two major categories of modification.

The first is prompt engineering, where the input to the model is modified to influence the result. The second if fine tuning via embeddings. Our current investigation indicates that prompt engineering is flexible but can be resource-intensive due to the large amount of tokens it requires. Fine-tuning via embedding vector encodings of additional content the model was not trained on is, by contrast, computationally efficient, but relatively expensive. We plan to initially adopt the heuristic that long-term knowledge should be embedded, whereas prompt engineering should be utilized for the short-term context and interactions.

4 CONTEXTUAL FACTORS IN HUMAN-ROBOT ENCOUNTERS

In this section, we elaborate upon the elements of context relevant to an interaction scenario in which a visitor makes incidental conversational contact with a mobile delivery robot equipped with an LLM-driven communication interface trained on information relevant to and gleaned from tours of the university campus where it makes its deliveries. Robots capable of performing both such delivery and tour-guiding tasks are currently in active development. Elements of this context are relevant as well to other aspects of the robot's behavior, such as social navigation. The intersection of contextualization required for conversational and navigational abilities is discussed further at the end of the paper.

4.1 Spatiotemporal Context

To interact capably with pedestrians, robots must modulate their behavior to the time and place of encounter. We will utilize our project's longitudinal, multimodal data collection to understand the spatiotemporal patterns on the campus and encode these in forms useful to conversational interaction but also functions like social navigation.

4.2 Interactional Context

To be perceived as credible conversational agents, robots must keep track of their relationships to those around them and the history of interactions with different people. This will require encoding those relationships into participant words, but also engineering appropriate expansion into those prompts so that, for instance, affect detection or gestural content might be included. In essence,

4.3 Role- or Task-based Context

Robots performing other tasks or duties who engage with InCoPsmust be able to modulate their responses in light of their existing responsibilities.

5 ENCODING CONTEXT FOR LONG-TERM MOBILE AUTONOMY

This section surveys the initial approaches we plan to take towards encoding context into forms tractable to LLMs.

5.1 Embedding Spatiotemporal Semantics

In the task of providing guided tours, knowledge of the names, history, and significance of the surrounding buildings is a core competency. We believe that encodings of this information represent

the most promising method of endowing robots with the capability of knowing about the locations to which some of them guide tour attendees, and around which others of them move in the course of carrying out other tasks such as delivery.

The following exchange demonstrates the potential role of spatial and temporal factors on conversational interactions with delivery robots.

> Delivery Robot: Excuse me, are you Julie? I have a delivery of hand sanitizer for Julie.
> Sally: No! My name is Sally. Hey, you can talk? Can you tell me where I am?
> Delivery Robot: You are on the campus of the University of Texas. What can I tell you about?
> Sally: What is up with that tower?
> Delivery Robot: Built in 1937, the Tower was originally designed to house the University's library. The collection eventually outgrew the space, and the building is now home primarily to University administration.
> You couldn't see it now because it's daylight out, but the Tower is often illuminated in burnt orange to celebrate events like sports victories or university graduation.
> Sally: There are a lot of quotes written on the building. Were they part of the original design or added later?
> Delivery Robot: Yes, the main inscription here on the front of the building was chosen by the building committee about two years before its completion.
> There's an interesting story about that, actually. Dr. William Battle initially proposed the quote be "The records of the past shall bring light and courage to them that come after." After that was received rather cooly by the committee members, Battle presented it with two other suggestions. One of the others was "*Cognoscetis Veritatem et veritas liberabit vos*" — Battle was a Classics scholar. The one that won out was the English version you see here, "Ye shall know the truth and the truth shall make you free." That's a quote from the New Testament, which was likely familiar to many at the time.
> Sally: This is so interesting. Tell me more!
> Delivery Robot: I am sorry, I have to continue with my delivery now, but if you wait at the orange sign to your left a guide robot will be leading a tour at the top of the hour.

5.2 Prompt Engineering to Encode Observations about Interlocutors

The fine-grained understanding of individual interlocutor interests is core to the individuation of tour details and itenerary. While admittedly advanced, a robot that remembered and attributed specific interests to tour participants would ideally be able to highlight facts about tour stops most likely to be interesting to attendees. For example:

> Tour Robot: Sally, you asked earlier about the architectural details of the Tower. If you look to our left as we head this way, you'll see more of the engraved quotations you were interested in. In this case, this building is the School of Architecture, but it was built around the same time. It's interesting that, although the details are classical, the inscriptions are in English, not Greek or Latin.
> ...
> Tour Robot: Alright, once everyone is here, we can begin.
> ...
> We're now in front of the Perry-Castañeda Library. No inscriptions on this one, Sally! Completed in 1977...

5.3 Potential Workflow for LLM Customization

A workflow to facilitate the development of customized LLMs using our project's longitudinal data as spatiotemporal contextualization would need to address at least the following:

- Encoding and updating semantic knowledge of the campus as it changes (e.g. when a building is renamed)
- Gathering and analyzing conversational interactions between robots and tour attendees for potential use in fine-tuning

In addition, while commercial services are likely to be used in initial work, the use of open source models and data sets alongside existing high performance computing resources such as **TACC!** (**TACC!**) will be key to sustainability and reproducibility.

6 DISCUSSION

The position we take at this point in our investigation is one of cautious and contingent optimism. Notwithstanding the substantial challenges specific to implementing LLMs in a long-term deployment of autonomous mobile robots, and concerns about their potential harms, we believe research in this area is clearly warranted. With careful and deliberate identification and mitigation of potential harms, effective contextualization, and means of adaptation and evolution to emergent or changing factors, LLMs could represent an important tool to help ensure that mobile autonomous robots have positive and equitable impacts on communities living and working with robots: communities which those LLMs help empower with regard to those robots through the contextual conversational interfaces they enable.

We present this work now to seek contributions from other researchers using LLMs in similar and different HRI research settings, constructive feedback on our reasoning, and inspirations for methods best suited to for harnessing the considerable capabilities of LLMs to contextual conversational interactions.

One area of particular interest to us is that overlaps of context relevant to conversation and navigation will render spatiotemporal and social awareness a core competency for autonomous mobile robots in long-term, community-embedded deployments. Future research should investigate the coordination of contextual encodings so that LLM embeddings, for instance, can either be transformed for use in social navigation tasks, or that some intermediate representation capable of generating the appropriate data structures for each be devised.

6.1 Potential Risks and Mitigation Strategies

The potential for LLMs to generate racist, sexist, or other harmful language is well-documented [1]. Hundt et al. [2] note that robots present a particularly concerning platform to utilize this technology, since their actuators mean that they might physically enact the biases encoded in model.

While the running example in this paper does not utilize LLM output for actuator control, the potential for overlapping relevant conversational and navigational context potentially could. This suggests that researchers must implement robust safeguards, testing, and auditing of robot-generated speech and any LLM-influenced actuation.

Several strategies hold promise for mitigating these risks. These include output filters, prompt engineering as a form of interpretive constraint, and potentially embedding-based corrections. Key to progress in this area will the development of ethically oriented benchmarks for specific LLM configurations and their integration into research workflows.

ACKNOWLEDGMENTS

This work was supported in part by Good Systems, a UT Grand Challenge, and NSF Award #2219236. The authors would like to thank Joel Fischer and other TAS Hub researchers for ongoing discussions that have informed this work.

REFERENCES

[1] Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? ☒. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (FAccT '21). Association for Computing Machinery, Virtual Event, Canada, (Mar. 2021), 610–623. ISBN: 9781450383097. DOI: 10.1145/3442188.3445922.

[2] Andrew Hundt, William Agnew, Vicky Zeng, Severin Kacianka, and Matthew Gombolay. 2022. Robots Enact Malignant Stereotypes. In *2022 ACM Conference on Fairness, Accountability, and Transparency* (FAccT '22). Association for Computing Machinery, Seoul, Republic of Korea, (June 2022), 743–756. ISBN: 9781450393522. DOI: 10.1145/3531146.3533138.

[3] Alec Radford et al. 2021. Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning* (Proceedings of Machine Learning Research). Marina Meila and Tong Zhang, (Eds.) Vol. 139. PMLR, 8748–8763. https://proceedings.mlr.press/v139/radford21a.html.

[4] Astrid Rosenthal-von der Pütten, David Sirkin, Anna Abrams, and Laura Platte. 2020. The Forgotten in HRI: Incidental Encounters with Robots in Public Spaces. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction* (HRI '20). Association for Computing Machinery, Cambridge, United Kingdom, (Mar. 2020), 656–657. ISBN: 9781450370578. DOI: 10.1145/3371382.3374852.