Ommie in the Home: A Non-Invasive Learning Model for Social Robotics

Grace Abawe[1]; Advisor Brian Scassellati[2]; Advisor Rajit Manohar[3]

***Abstract*— Protecting user privacy is essential to keep in mind when designing robots meant for in-home deployment because of the presence of potentially sensitive and user-identifiable information within the environment. However, when it comes to Human-Robot Interaction (HRI), creating smart, sociable robots is also an essential but difficult task. This work explores the intersection of these two challenges, ultimately proposing an approach that produces a system to predict the user's intent to interact with a robot without the usage of sensitive data. The system leverages a Gated Recurrent Unit (GRU) Network to make predictions anticipating the intent to interact based on the more sparse information captured by a wave radar sensor, as opposed to popular methods that use cameras. This system is situated within the context of the Ommie robot, an anxiety reduction robot meant for in-home use. A novel dataset is produced, consisting of nearly 100 trials of participants deciding to interact or not to interact with Ommie and includes data captured by radar, microphone, RGB camera, depth camera, and capacitive touch sensors. Results indicate that radar data does yield accurate results for predicting intent to interact and thus can be an equally useful sensor for contributing to intelligent robot behaviors, whilst collecting far less invasive data.**

INTRODUCTION

Many people are sensitive to being observed and monitored in the comfort of their home. At the same time, devices and robots that collect data within the home are becoming increasingly popular. Still, consumers are rightfully worried about in-home technological surveillance, which has been a concern since the advent of always-on voice assistants like Alexa. Yet, Amazon themselves have stated that Alexa does not record audio if it hasn't been voice-activated and that they do not sell users' personal data [1]. And while robots like the Roomba do collect mapping and navigation information, this data is necessary for mapping out rooms for vacuuming [2]. While some consumers aren't uncomfortable with the idea of a robot in their home that freely uses a camera, microphone, or is connected to the Internet of Things (IoT), many others may prefer to prioritize privacy within their homes. Though these robots may be collecting data to better perform their jobs, it may not be worth the cost of sacrificing user privacy. This principle must be kept in mind when designing robots meant for in-home use.

However, some experts in the field of robotics argue that this could be to the detriment of potential technological progressions [3]. Data collection by robots is used not only for robot functionality but also to customize robots to user preference and to provide performance feedback to engineers. Social robotics in particular hinges on established trust and companionship between the robot and the human as the most important factor for humans to want to continue interacting with the robot [4].

Interacting with humans can take place in personal settings, such as the home, in schools, or in healthcare facilities. In these environments, a robot should not freely collect all available data, even for the sake of facilitating positive social interactions. Thus, social robots must balance the trade-off between collecting more data, which allows for more complex models that accurately predict human intentions, with limiting data collection, which protects privacy in sensitive environments but can potentially yield less accurate results.

The question is, then, how do engineers design robots to maximize helpfulness and personability without invasively collecting user data? Additionally, we must also define invasive data collection, both in terms of the type of sensor data collected, what environmental data is available, and when it is collected (as in, is the system considered "on" or "off" at the time of collection). In this paper, the problem space is constricted so as to only be concerned with the types of sensors utilized, the types of data collected by these sensors, and how user-identifiable that data is. I hypothesize that the data produced by lower-level sensors can still be relatively useful for certain tasks regarding human-robot interaction. Ultimately, I deliver a system that can be deployed on an in-home robot and is able to leverage non-sensitive data to predict human intent to interact.

BACKGROUND

In-home robots are designed for many uses, including social and therapeutic ones [5], [6]. Ommie is a robot designed to reduce anxiety by helping guide users through deep-breathing exercises [7]. Users interact with Ommie by putting their hands on Ommie's body, which expands and contracts in time with the breathing exercise. Ommie is equipped with a capacitive touch sensor, digital eye screens, motors, a speaker, silicone skin, and a RaspberryPi, as seen in Figure 1. Great care has been taken to ensure Ommie's design has been optimized to be calming, engaging, and successful in instructing deep-breathing exercises. Effectively, Ommie should function similarly to a pet in its companionship but also like an assistant or a coach in its ability to be helpful to humans. Still, there are things that can be added to Ommie's existing design to maximize Ommie's helpfulness

[1]Student of Electrical Engineering nad Computer Science, Yale University.

[2]Brian Scassellati is with the Faculty of Computer Science, Yale University.

[3]Rajit Manohar is with the Faculty Electrical and Computer Engineering, Yale University.

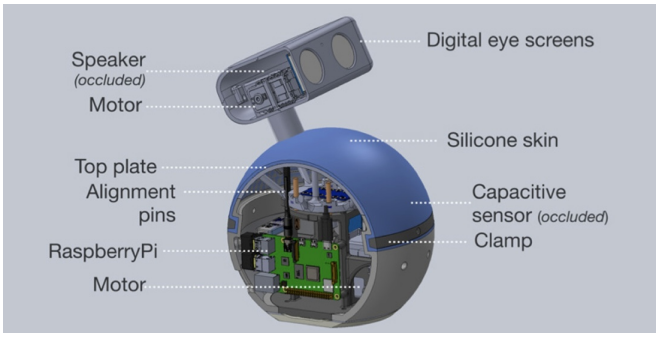

Fig. 1. System Design for the Original Ommie Robot [7]

and sociability for each individual user, while still being mindful of user privacy.

If Ommie were to collect data 24/7, it is almost certain that Ommie would be able to better customize its usefulness to the user. However, if a future Ommie or Ommie-variant were to ever be deployed in a home or some other privacy-preserving space (such as a psychiatrist's office), Ommie needs to be able to retain these important, sociable characteristics while being minimally invasive in its data collection. Any added sensors must be selectively chosen so that captured data is non-invasive.

Beyond sensor selection, another problem is whether the data collected can be leveraged by Ommie to make it a more helpful anxiety-reduction robot. If it is concluded that data from a low-level sensor like a passive infrared motion detector is much less useful than data collected from something like a camera for facilitating certain robotic behaviors, then at least users can be aware of the concessions they are making when deciding whether to purchase an in-home robot equipped with certain sensors. Users should be able to make informed decisions about the robots within their home, which is why it is essential to know the capabilities of a robot given the sensors it is equipped with.

Related Work

Prior work has investigated privacy in Human-Robot Interaction (HRI), including in-home robot deployment [3], [8]–[13]. However, much of this research investigates the proper usage and disposal of potentially sensitive data. While ensuring that user data isn't saved long-term or used inappropriately (like being sold to third-parties) is extremely important, this problem can also be avoided altogether by limiting the collection of sensitive data.

There is also a pool of research investigating sensor usage in the context of HRI for the purposes of predicting human behavior [14]–[19]. Much of these methods incorporate the employment of methods such as gaze, gesture, and speech detection that are then used to train Neural Networks. Still other research has investigated limited sensor and data usage to similarly achieve various complex goals in robotics [20]–[22]. However, work specifically in preserving privacy within social robotics based on limited sensor usage is relatively sparse. This is the research area this work falls within.

Methodology

Much of this work was dedicated to ensuring the proper approach to solving these problems: maintaining the principles of preserving user privacy, upholding research ethics, and high reproducibility of experiments. As a result, the methodology used to obtain data is thoroughly documented.

Approach

To further restrict the problem domain, I focus on one helpful behavior Ommie should be able to perform when deployed in a home. To seem more alive, Ommie should be able to sense and react accordingly when a person or people are trying to interact with it. Specifically, Ommie should recognize a person's intent to interact and exhibit some behavior (i.e. eye contact or a greeting noise) to show that it recognizes this intent. Intent to interact is the concept that if a person has the *intention* of interacting, then in the near future they will indeed interact with the robot [23]. Just as important as being able to anticipate an interaction is knowing how long the gap between the initial intent and the actual interaction is. A Bohus and Horvitz study found that the intent to interact occurs on average four seconds before the actual interaction [24]. Thus, I also label the intent to interact as four seconds before the actual interaction. A smart robot should be able to respond to a person who approaches it with the intent to interact, so incorporating this ability would make Ommie more sociable overall.

Recognizing the intent to interact has been studied and successfully done before, such as in Thompson's 2024 study [23], though usually using sensors like cameras. Thus, I similarly explored whether Ommie could predict a person's intent to interact by using sensors other than an RGB camera since most users concerned with privacy likely wouldn't approve of using facial detection methods. Instead, I studied whether Ommie could predict the intent to interact using lower-level sensors. Additionally, I employed an RGB and depth camera as baselines for comparison. Using the data collected from these sensors as inputs, I trained a Gated Recurrent Unit Recurrent Neural Network (GRU) model to predict a person's intent to interact. The prediction is a boolean output. Either a user *does* intend to interact, or *does not* intend to interact, as can be seen in Figure 2. Then, Ommie can use the model's output prediction to activate some sort of basic greeting, such as waking up and blinking at the approaching person, or saying "Hello!".

Sensors

This section describes the details of each sensor used in the experiments and why they might or might not be the preferable choice for users concerned about privacy. The entire hardware flow setup can be observed in Figure 3.

Wave Radar Sensor[1]

This radar sensor emits carrier waves that are reflected by objects. The time elapsed between transmission and

[1]Acconeer A121 Pulsed Coherent Radar (PCR)

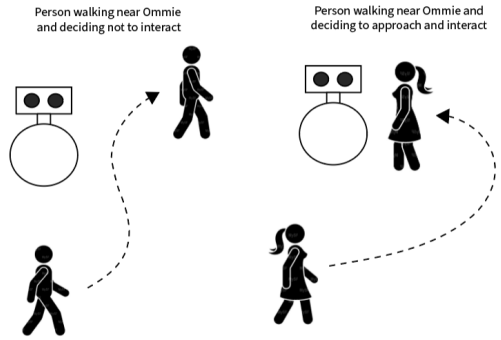

Fig. 2. Two separate individuals walking within proximity of Ommie. The person on the left walks near Ommie but ultimately decides not to interact. The person on the right walks up to Ommie and does decide to interact.

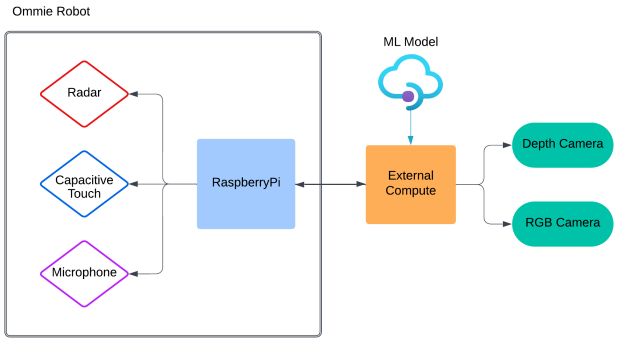

Fig. 3. Ommie Hardware Flowchart

reception of the signal is used to calculate the distance to the object.

This sensor can detect distance, speed, and general motion. It can also recognize breathing, hand motions, vibrations, and detect presence. For this study, the radar sensor was used to detect presence, which is necessary for predicting intent to interact. Presence is determined from intra-frame presence (detecting fast movements inside frames) and inter-frame presence (detecting slow movements between frames) of sub-sweeps. An example of plotted presence detection data can be observed in Figure 4.

Radar sensors can predict presence and recognize specific hand gestures, but cannot gather any identifying features of people (or objects, for that matter). Depending on the sweep range, the values collected could also be turned into a depth image vector, similar to but less detailed than the ones collected by a depth camera, as described below. This is great for maintaining privacy, but might also yield lower accuracy.

Multi-directional Microphone[2]

The microphone used has four microphone arrays and can capture sound, decibel level, direction of arrival, and detect voice activity.

For our purposes, the audio file itself was not used, just

[2]ReSpeaker Mic Array v2.0

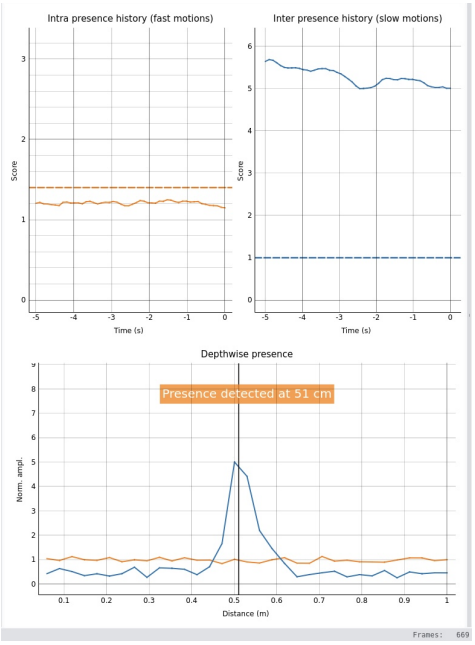

Fig. 4. Radar Sensor Presence Detection Plot. In this instance, the values of the intra-presence and inter-presence scores led to the conclusion that a presence was detected at a distance of 51cm.

the decibel, direction, and voice activity values extracted during runtime, which contain no identifiable features. Additionally, even if further processing of actual audio data was done in real time so that other helpful features could be extracted or for keyword recognition (phrases like “Hi, Ommie”), this could potentially be considered non-invasive so long as audio isn’t saved, only processed in real-time.

Capacitive Touch Sensor (CapTouch)[3]

The touch sensor detects capacitive loads on touch contacts (i.e., the sensor is touched). The sensor utilized was a 5-pad touch sensor, which can distinguish which pad was touched. For our purposes, all pads were treated the same since the area the sensor was touched did not matter.

Camera[4]

Data collected by the camera was used as a baseline for comparison to the radar. The depth image captures less distinguishing features compared to RGB images, but more distinguishing than radar. The RGB image was used as a baseline for both the depth image and the radar. It does not preserve user privacy. Refer to Figure 5 for an example of how the depth image differs from the RGB image in the context of capturing user-identifiable characteristics.

Other Setup

RaspberryPi 5

The Raspberry Pi 5 microcontroller, which is used in the actual Ommie robot, ran the Python scripts for the sensors during data collection. The Raspberry Pi ran the

[3]Standalone 5-Pad Capacitive Touch Sensor Breakout - AT42QT1070
[4]Intel® RealSense™ Depth Camera D435i

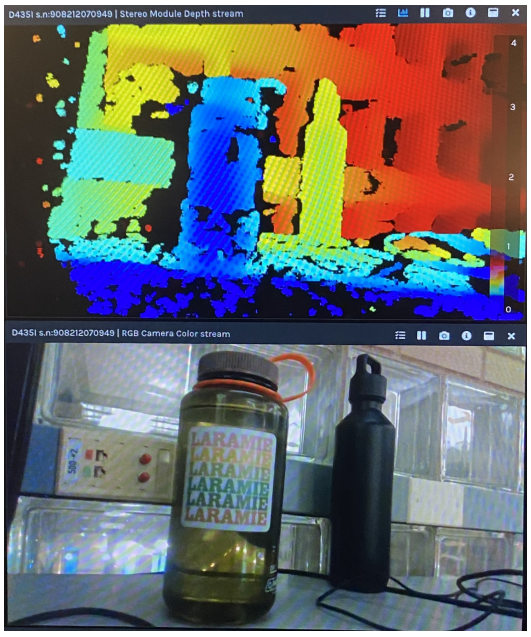

Fig. 5. Depth Image versus RGB Image of Two Water Bottles

microphone, capacitive touch, and radar.

External Compute

Because the RealSense camera is not compatible with RaspberryPi, it was set up on a separate MINISFORUM mini computer and ran in parallel with the RaspberryPi. The GRU model was also not run in realtime on either the RaspberryPi or the mini computer.

Speaker

A speaker was utilized to output a sound when Ommie was successfully interacted with so that the participant had concrete feedback that Ommie acknowledged the interaction. The speaker played a random sound from a library of Ommie sounds, such as a giggle or a sigh, when the capacitive touch sensor was touched.

Model Ommie

A foam, life-sized model of the Ommie robot was used for data collection to avoid spending further time integrating all new sensors with the preexisting Ommie robot unnecessarily (which would also require using ROS). Using a model was also necessary so that sensor placement would remain realistic and people still had a "robot" to look at and interact with.

On the foam model, the capacitive touch sensor was attached to the foam body's midsection (so that interacting with Ommie required touching the body), the radar sensor was right above the capacitive touch sensor, and the microphone was above both, positioned horizontally under the chin to collect 360 degrees of sound input. The camera was placed on Ommie's head, closest to his eyes. Lastly, a speaker was placed on the table next to Ommie. The model was kept in the same position throughout all trials to maintain consistency. Refer to Figure 6 for the foam model setup.

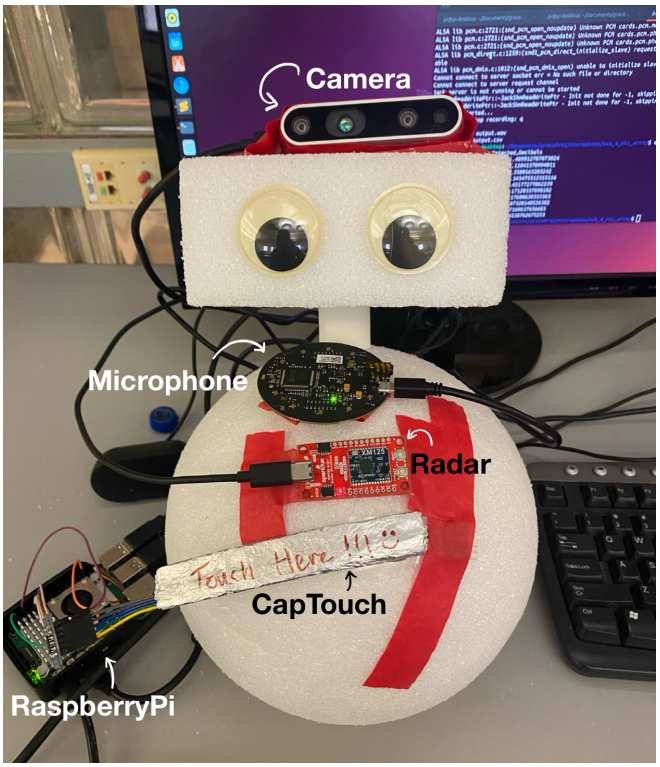

Fig. 6. Foam Ommie Model Design and Setup

Feature Engineering

Due to the differences in size of the input data, embeddings of the camera outputs, including both the RGB image and the depth image, were created. These embeddings were made during the data preprocessing step before being input into the ML model. This is an important data compression step to shrink the high dimension inputs to a lower dimension that is more similar to the other sensor data. Otherwise, because of the different modalities, some features would dominate due to scale differences. The videos are processed using a ResNet Convolutional Neural Network (CNN) with 18 layers to create the embeddings. Further feature engineering was also performed on the radar dataset, where microphone values were padded to match the scale of radar data at each timestep.

Machine Learning Model

A Gated Recurrent Unit is a type of Recurrent Neural Network (RNN) that is the simpler alternative to Long Short-Term Memory (LSTM) networks. A GRU is the ideal model for this multimodal dataset because it can process sequential data, such as the radar and audio input, and can capture the temporal component inherent to intent prediction. It can also capture long-term dependencies, which can be helpful for detecting subtle changes in behavior patterns over time. Figure 7 shows the pipeline of preprocessing input data that is then fed to the model for a prediction to be output.

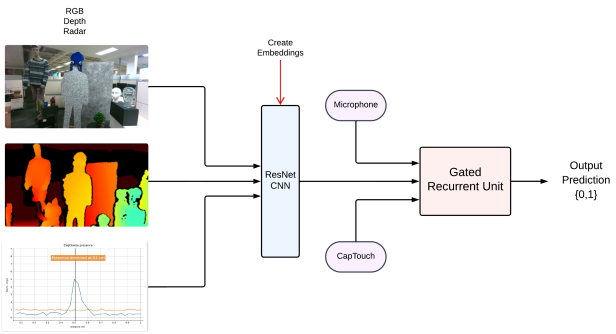

Fig. 7. Feature Embeddings and Gated Recurrent Unit Pipeline

GRUs use gating mechanisms that selectively update the hidden state at each timestep. GRUs use three gates: the update gate, reset gate, and current memory gate. The update gate determines how much of the past knowledge should be passed along into the future. The reset gate determines how much past knowledge to forget. Lastly, the current memory gate is incorporated into the reset gate to introduce non-linearity into the input. GRUs take the current input and the previous hidden state to calculate the values of the three different gates. The output is then the prediction regarding intent to interact.

The parameters of the final GRU included 2 layers, a hidden size of 64 (meaning input is processed into a 64-dimensional hidden state), and a dropout rate of 0.2 between layers. The output prediction was a binary classification.

Data Collection

The dataset consists of nearly 100 trials of participants either interacting with or ignoring Ommie. Each trial varies in length but are often around 10-20 seconds long. In each, a participant (or in a handful of cases, two participants) either walked past Ommie or approached Ommie, as in Figure 2. Of those who approached it, some then decided to interact with Ommie by stepping up to touch the CapTouch sensor on Ommie's belly (which triggered Ommie to make a random sound of acknowledgement), once or multiple times[5] before walking away. This setup was designed to mimic how users might interact with Ommie in a home or how people might approach a robot in public. In real life, a user in the home may happen to be in the same room as Ommie and may even go near Ommie if they are doing something in the vicinity of Ommie, but may not actually intend to interact with Ommie. Similarly, people in public sometimes choose to interact with a robot if they are curious about it while others may choose not to.

In order to verify the interaction, the world timestamp that the capacitive touch sensor was first touched was saved (and any subsequent touches ignored since the interaction had already begun) when a participant decided to touch Ommie. Additionally, when the script that controlled the sensors was run, the initial world timestamp was also saved for reference. Because the camera was not run on the RaspberryPi, its initial world timestamp was also saved, which was usually a few deciseconds off from the other sensors. As a preprocessing step, all of the initial world timestamps were matched up so that the data was correctly lined up.

[5]For the purposes of maintaining participant engagement should they like to hear Ommie make another sound, sounds continued to play if the CapTouch was pressed more than once, but only the timestamp of the first touch would be recorded as the beginning of the interaction.

Results

The dataset was split 80/10/10 between training, validation, and testing. The loss function used was Cross-Entropy loss and the Adam Optimizer was set to an initial value of 0.001. Gradients were clipped to a maximum of 5 to prevent exploding gradient issues, and the number of epochs used for training was 10.

The GRU made predictions on the intent to interact, outputting either 0 or 1 at each timestep. Table I shows the results of the predictions of the GRU model based on the input radar and microphone data. It can be observed that the model predicted remarkably well. One concern is whether this was due to overfitting from hyperparameter tuning, but Figure 8, which compares training with validation loss, shows that there was not significant divergence between training and validation loss, which is usually indicative of overfitting. Figure 8 also depicts the steady decrease in loss at each epoch, which implies that the model was optimizing weights effectively during training to find patterns. At the first epoch, the training and validation loss was 0.7499 and 0.7103, respectively, by the 5th epoch 0.4680 and 0.4246, and by the 10th epoch 0.2215 and 0.1812. The best method for confirming whether or not the model overfits would be a physical robot demonstration that uses the model to predict intent to interact in real-time, which is worth looking into as a follow-up.

In the future, these results could also be compared to the depth and RGB camera data as baselines. I hypothesize that these baselines would perform comparably well because it produces even more complex data before being compressed into embeddings, but that the GRU hyperparameters might require different tuning to fit different data representations.

Conclusion

Based on the results of the model predictions, the use of a radar and microphone for predicting intent to interact indeed produces comparable results to other, higher-level sensors that can capture more features. Thus, a robot system that facilitates more human-like social behavior was successfully created. It is also observable that regardless of if the model does overfit, this method is still useful for predicting the intent to interact. For users who value the privacy that a radar provides, this is a very feasible option. Before fine tuning the model, it was observed that the system tended to provide false positives more often than false negatives, meaning Ommie would incorrectly greet a user even if they were not trying to interact with Ommie. Even this can be interpreted as more helpful than a system that produces too many false negatives, since among the users who initially

TABLE I
RESULTS OF MODEL PREDICTIONS.

Sensor	Accuracy	Precision	Recall	F1 Score
Radar + Microphone	0.9987	1.0000	0.9987	0.9993

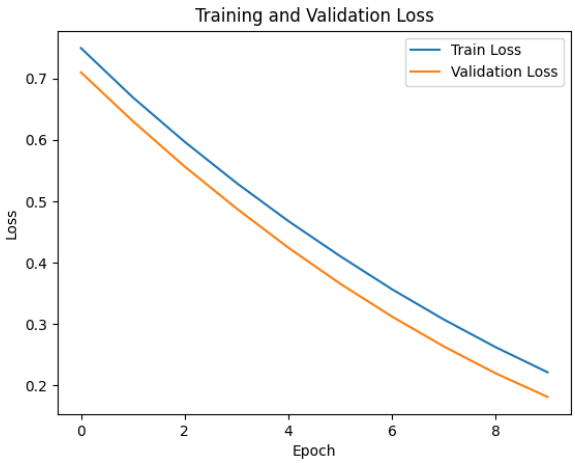

Fig. 8. Model Training vs. Validation Loss over Ten Epochs

approached the robot but ended up walking away early before touching the robot, a greeting from Ommie could potentially influence them to go through with the interaction.

Additional experiments can be conducted to gather more data and further validate results. Likewise, other models such as a Random Forest Model could also be worth investigating, so as to compare the performances. If a relatively simpler model produces comparably good results, migrating to the simpler model would be beneficial. If generalizability is a concern, another strategy could be to generate lots of data points using a simulation instead of collecting data in the physical world. Data generated in simulation would provide the flexibility of more data to work with, but would also come with the cost of having to deal with the sim-to-real gap, which is the transferability of results from simulation into real-life (since simulations will always produce better results due to being less noisy).

Overall, further work investigating privacy in the context of specific sensor usage and in-home deployable robotics is necessary not only from a scientific standpoint, but also from a social one. These results are promising in that future work can lead to further breakthroughs in utilizing lower-level sensors to produce similar results to higher-level ones for predicting intent-to-interact, as well as other social behaviors. Thus, future researchers should remain optimistic in regards to this topic.

FUTURE WORK

Firstly, it is worth looking into other specific low-level sensors that could similarly be used to observe how they might perform in comparison to the radar sensor that was selected. This includes sensors like an infrared motion detector, lidar, or even a light sensor, as well as any combination of low-level sensors. These may yield different results as well as have their own ratings in how much privacy they preserve relative to radars. Additionally, further trials can be run to create a larger, more robust dataset to further validate results.

In the context of Ommie's main function being a deep-breathing robot for in-home use, another useful user-interaction to implement in the future would be to have Ommie remind the user to do their breathing exercises. The difficulty is that Ommie should have a strategy for picking the optimal time to remind the user, as well as to adapt to the user when it comes to *how* Ommie should sound when giving a reminder. As for picking the optimal time, this is tricky because it becomes a problem of categorizing contexts and user preference. Ommie must detect the current context, as well as decide whether the context is appropriate for giving the user a reminder.

One scenario might be in the morning when the user first walks into the room Ommie is in. While the user might have time to do an exercise, they might be planning on doing their exercise later in the day. Contrast this with the scenario where it is the end of the day and the user has yet to do their exercise, but is also deeply entrenched in their work. In which scenario should Ommie try to remind the user to do their exercise? Of course, the answer also depends on user preference, since one user might appreciate the reminder to take a break, and another might be irritated by the sudden interruption.

How the reminder sounds is important as well. Things like wording, tone, and firmness preferences differ widely from user to user. Some users may prefer for their reminders to be gentle, while others may want a more stern approach. Roshni Kaushik and Reid Simmons demonstrate the importance of keeping user preferences in mind in their research involving an exercise coach robot that provides user feedback. Kaushik and Simmons found that robot feedback can improve human performance, but that users responded differently to "encouraging" styled coaching, as opposed to "firm" styled [25]. With this in mind, further research must be conducted to evaluate the general approaches users prefer the most and how Ommie should classify its user to predict what approach they would be most responsive to.

Lastly, the system can be integrated into the actual Ommie robot to be used in real-life. Before then, a physical robot demonstration should be done to confirm usefulness and consistency of the system and to troubleshoot potential problems that may arise. Integrating this system into a real robot is a long-term goal, so that the larger pool of available robotic consumer products includes systems with more emphasis on privacy in the future. To reiterate, much more important work can, and should be done in this research area.

ACKNOWLEDGEMENTS

This work was generously supported by the Timothy Dwight Mellon Senior Forum Research Grant and I would like to thank those at the TD Head of College Office for their generous funding.

I want to express my gratitude to my advisor Brian Scassellati and second reader Rajit Manohar for their expert guidance. For her support and guidance, I also thank my graduate mentor Fern Limprayoon, as well as Kayla Matheus and Sydney Thompson. Lastly, I'd like to thank Andy Cheng, who has been instrumental in sharing his knowledge on hardware setup. Special thanks to the Yale Department of Computer Science and the Department of Electrical Engineering, for providing opportunities for academic growth.

REFERENCES

[1] A. Staff, "Busting the 5 biggest myths about amazon's alexa," Jun 2022. [Online]. Available: https://www.aboutamazon.com/news/devices/busting-the-5-biggest-myths-about-amazons-alexa

[2] Nov 2023. [Online]. Available: https://homesupport.irobot.com/s/article/964

[3] S. Chatterjee, R. Chaudhuri, and D. Vrontis, "Usage intention of social robots for domestic purpose: From security, privacy, and legal perspectives," *Information Systems Frontiers*, vol. 26, pp. 121–136, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:239197387

[4] P. Kellmeyer, O. Mueller, R. Feingold-Polak, and S. Levy-Tzedek, "Social robots in rehabilitation: A question of trust," *Science Robotics*, vol. 3, no. 21, p. eaat1587, 2018. [Online]. Available: https://www.science.org/doi/abs/10.1126/scirobotics.aat1587

[5] S. Sabanovic, C. Bennett, W.-l. Chang, and L. Huber, "Paro robot affects diverse interaction modalities in group sensory therapy for older adults with dementia," *IEEE ... International Conference on Rehabilitation Robotics : [proceedings]*, vol. 2013, pp. 1–6, 06 2013.

[6] A. A. Scoglio, E. D. Reilly, J. A. Gorman, and C. E. Drebing, "Use of social robots in mental health and well-being research: Systematic review," *J Med Internet Res*, vol. 21, no. 7, p. e13322, Jul 2019. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/31342908

[7] K. Matheus, M. Vázquez, and B. Scassellati, "A social robot for anxiety reduction via deep breathing," *IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, vol. 31, pp. 89–94, 2022.

[8] S. Chatterjee, R. Chaudhuri, and D. Vrontis, "Acceptance of social robot and its challenges: From privacy calculus perspectives," *Technological Forecasting and Social Change*, vol. 196, p. 122862, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0040162523005474

[9] M. Rueben, A. M. Aroyo, C. Lutz, J. Schmölz, P. Van Cleynenbreugel, A. Corti, S. Agrawal, and W. D. Smart, "Themes and research directions in privacy-sensitive robotics," in *2018 IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO)*, 2018, pp. 77–84.

[10] T. Schulz and J. Herstad, "Walking away from the robot: negotiating privacy with a robot," in *Proceedings of the 31st British Computer Society Human Computer Interaction Conference*, ser. HCI '17. Swindon, GBR: BCS Learning & Development Ltd., 2017. [Online]. Available: https://doi.org/10.14236/ewic/HCI2017.83

[11] C. Lutz and A. Tamò Larrieux, "Do privacy concerns about social robots affect use intentions? evidence from an experimental vignette study," *Frontiers in Robotics and AI*, vol. 8, p. Article 627958, 04 2021.

[12] C. Lutz, M. Schöttler, and C. Hoffmann, "The privacy implications of social robots: Scoping review and expert interviews," *Mobile Media & Communication*, vol. 7, pp. 412–434, 09 2019.

[13] P. Liu, D. F. Glas, T. Kanda, and H. Ishiguro, "Data-driven hri: Learning social behaviors by example from human–human interaction," *Trans. Rob.*, vol. 32, no. 4, p. 988–1008, Aug. 2016. [Online]. Available: https://doi.org/10.1109/TRO.2016.2588880

[14] A. A. Abdelrahman, D. Strazdas, A. Khalifa, J. Hintz, T. Hempel, and A. Al-Hamadi, "Multi-modal engagement prediction in multi-person human-robot interaction," *IEEE Access*, vol. PP, pp. 1–1, 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:249656534

[15] G. S. Jam, J. Rhim, and A. Lim, "Developing a data-driven categorical taxonomy of emotional expressions in real world human robot interactions," *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:232136593

[16] D. Strazdas, J. Hintz, A. Khalifa, A. A. Abdelrahman, T. Hempel, and A. Al-Hamadi, "Robot system assistant (rosa): Towards intuitive multi-modal and multi-device human-robot interaction," *Sensors (Basel, Switzerland)*, vol. 22, 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:246318849

[17] A. B. Youssef, G. Varni, S. Essid, and C. Clavel, "On-the-fly detection of user engagement decrease in spontaneous human–robot interaction using recurrent and deep neural networks," *International Journal of Social Robotics*, vol. 11, pp. 815 – 828, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:204350395

[18] A. Khalifa, A. A. Abdelrahman, D. Strazdas, J. Hintz, T. Hempel, and A. Al-Hamadi, "Face recognition and tracking framework

for human–robot interaction," *Applied Sciences*, 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:249248954

[19] Z. Zhang, J. Zheng, and N. Magnenat-Thalmann, "Engagement intention estimation in multiparty human-robot interaction," *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*, pp. 117–122, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:237296309

[20] K. Matheus, E. Mamantov, M. Vázquez, and B. Scassellati, "Deep breathing phase classification with a social robot for mental health," in *Proceedings of the 25th International Conference on Multimodal Interaction*, ser. ICMI '23. New York, NY, USA: Association for Computing Machinery, 2023, p. 153–162. [Online]. Available: https://doi.org/10.1145/3577190.3614173

[21] V. Mollyn, R. Arakawa, M. Goel, C. Harrison, and K. Ahuja, "Imuposer: Full-body pose estimation using imus in phones, watches, and earbuds," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, ser. CHI '23. ACM, Apr. 2023, p. 1–12. [Online]. Available: http://dx.doi.org/10.1145/3544548.3581392

[22] W. Bainbridge, S. Nozawa, R. Ueda, K. Okada, and M. Inaba, "A methodological outline and utility assessment of sensor-based biosignal measurement in human-robot interaction: A system for determining correlations between robot sensor data and subjective human data in hri," *International Journal of Social Robotics*, vol. 4, 08 2012.

[23] S. Thompson, A. Narcomey, A. Lew, and M. Vázquez, "Shutter: A low-cost and flexible social robot platform for in-the-wild deployments," in *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '24. New York, NY, USA: Association for Computing Machinery, 2024, p. 94–96. [Online]. Available: https://doi.org/10.1145/3610978.3641090

[24] D. Bohus and E. Horvitz, "Learning to predict engagement with a spoken dialog system in open-world settings," in *Proceedings of the SIGDIAL 2009 Conference: The 10th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, ser. SIGDIAL '09. USA: Association for Computational Linguistics, 2009, p. 244–252.

[25] R. Kaushik and R. G. Simmons, "Effects of feedback styles on performance and preference for an exercise coach," *2024 33rd IEEE International Conference on Robot and Human Interactive Communication (ROMAN)*, pp. 1516–1523, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:273694778