Check for updates

Data-Driven Natural Behavior Model Design with Large Language Models for Robotic-Animal Assisted Interventions (RAAI)

Keuntae Kim
Department of Biomedical Engineering
The George Washington University
Washington, D.C., United States
keuntae.kim@gwu.edu

Chung Hyuk Park
Department of Biomedical Engineering
The George Washington University
Washington, D.C., United States
chpark@gwu.edu

***Abstract*—Animal assisted intervention has been one of the effective natural therapeutic approaches, especially for individuals with autism. To increase the accessibility and reduce extra burden of care for the animals, Robotic-Animal Assisted Interventions (RAAI) has been proposed. However, the lack of natural behaviors is one of the key factors in limiting the feasibility with the current technology. This late-breaking report aims to build natural behavior models with data-driven approach, utilizing latest development of large language models (LLMs) to effectively analyze the data and build natural language-based models. Due to the reliance on LLM in this study, a key limitation is the lack of continuity in understanding. Frame images, as static representations, may not fully capture temporal dynamics. Future studies could address this limitation by integrating 3D-pose analysis, which would improve both continuity and contextual understanding.**

***Index Terms*—Autism; Neurodivergent; Robotic-Animal Assisted Intervention; Robotic Dog; Companion Robot**

I. Introduction

The prevalence of Autism* has been steadily increasing worldwide. According to recent studies, approximately 27.6 per 1,000 children are diagnosed with autism, highlighting its growing significance as a public health concern that warrants further attention [2].

Various therapies and interventions have been developed over time for autism, among which Animal-Assisted Intervention (AAI) has emerged as a notable approach. AAI has been widely acknowledged for its benefits in enhancing social interaction abilities, and emotional and cognitive functioning, as well as reducing problematic behaviors, particularly among autistic individuals and other neurodivergent conditions [3], [4]. However, AAI typically relies on "live" animals, which may pose challenges for individuals with animal allergies or other types of accessibility limitations. These barriers can prevent certain individuals from benefiting fully from AAI, even when it could be an optimal therapeutic approach for their needs. Robots, particularly robotic animals, present a promising alternative by offering programmable, predictable, and highly effective interaction tools [5]. They can effectively address the limitations of live animal interactions while replicating many of their therapeutic benefits. To increase the accessibility and reduce extra burden of care for the animals, this late-breaking report aims to set foundational knowledge for Robotic-Animal Assisted Interventions (RAAI).

Previous studies have demonstrated the therapeutic potential of robotic dogs as a method for canine-assisted intervention [6]. However, the lack of natural behaviors is one of the key factors in limiting the feasibility with the current technology [7]. This study, thus, seeks to build on these findings by addressing the limitations of earlier work and enhancing their effectiveness. Furthermore, it will explore the feasibility of this approach in real-world interaction environments.

II. Background and Related Works

A. *Autism (Autism Spectrum Disorder)*

Autism is a neurodevelopmental condition characterized by challenges in social communication, restricted and repetitive behaviors or interests [8]. Autism is highly heterogeneous, characterized by a broad spectrum of abilities and challenges, underscoring the necessity for personalized interventions [6], [9], [10].

Despite challenges in social interaction, autistic individuals often express a strong desire to connect with others, albeit in unique ways. Neurodiversity paradigms emphasize viewing autism as a natural variation in human development rather than a deficit [10]–[12]. Interventions should adopt this perspective by prioritizing the enhancement of quality of life and the promotion of meaningful social interactions, as well as fostering emotional and behavioral responses, rather than aiming to conform behaviors to normative standards.

B. *Animal-Assisted Intervention*

Animal-Assisted Intervention (AAI) has gained recognition for its potential to improve social, emotional, and cognitive

*We do recognize the clinical term of Autism Spectrum Disorder (ASD) but at the same time honor the recent discussion on the use of language in autism research [1]. Thus, we adopt a preferred (i.e., "autistic" vs. "with autism") and a less polarizing (i.e., "on the autism spectrum") terminology when referring to individuals with autism.

979-8-3503-7893-1/25/$31.00 ©2025 IEEE

skills in autistic individuals. Studies have shown that interaction with therapy animals, particularly dogs, can reduce anxiety, enhance social communication, and provide emotional comfort [13], [14]. Dogs are often favored for AAI due to their highly social nature and ability to respond to verbal and non-verbal cues, creating a feedback loop that encourages participants to engage meaningfully [4], [15].

Autistic children, in particular, benefit from interactions with therapy animals, as they often exhibit increased playfulness and focus in the presence of a therapy dog [3]. Therapy animals can serve as social bridges, enabling individuals to apply social skills developed through interactions with the animals to their interactions with real humans [11]. However, AAI may present practical challenges, such as limited accessibility, animal allergies, or the requirement for trained animals. As a result, robotic alternatives could provide an effective solution for individuals who face difficulties accessing traditional AAI.

C. *Animal Natural Behavior Learning*

Silva et al. (2019) found that living dogs are more effective than robotic dogs in enhancing social communication skills in children, whereas no significant difference was observed for adults [16]. This suggests that, for children, robotic dogs designed to closely mimic the realistic appearance and behavior of living dogs could have a similarly significant impact and be effectively personalized to meet the unique needs of autistic children.

In addition, Thodberg et al. (2016) found that interactions between older adults and robotic animals gradually decreased over time, while interactions with real animals remained consistent [17]. Nevertheless, Chang et al. (2020) suggested that advancements in robotic animal technology could make them increasingly comparable to living animals, potentially yielding equivalent or even superior outcomes in the near future [18]. Thus, continued advancements in robotic animal technology have the potential to enhance the quality and frequency of interactions between older adults and robotic animals in the near future.

Replicating the natural behaviors of real dogs in robotic dogs [19], [20] remains a significant challenge, primarily due to the limited availability of vision-based data. Collecting such data in lab-like settings on a large scale risks generating behaviors that may no longer be truly “natural.” However, the ongoing growth of online data offers a promising resource for developing models capable of accurately learning and replicating natural behaviors, provided that copyright restrictions are appropriately addressed.

III. Method

A. *Data (Videos)*

To replicate the natural behaviors of living dogs in robotic systems, this study utilized a dataset of video recordings capturing real dogs’ behaviors in everyday environments, specifically for training machine learning models on dog pose estimation. These videos were sourced from a publicly available GitHub repository [21], where the author provided them for neural network training purposes. The videos, featuring the author’s dog in both outdoor and indoor settings, capture various poses from multiple angles during interactions between the author and the dog. Designed explicitly for model training, the video dataset offers naturalistic and clean data that surpasses many other online sources not collected in controlled lab environments. Moreover, the consistency in camera angles and stability makes the dataset highly suitable for training purposes and video analysis.

To analyze dogs’ behaviors in a more naturalistic environment, we also included and utilized a YouTube video in this study. The video, recorded by a Ring home security camera, captures a dog’s activities when its owners are not at home [22]. As a home camera recording, it provides insights into natural behaviors exhibited by the dog without “direct” human interference. We hypothesized that this video could yield more authentic behavioral keywords compared to those derived from the GitHub video dataset, which was originally created for training neural networks as mentioned earlier.

The dataset serves as the foundation for developing models to accurately learn dog behaviors. To ensure quality and relevance, the videos had to meet specific criteria: first, they needed to depict natural, everyday environments such as homes, parks, or other familiar settings. Second, the videos had to capture a diverse range of behaviors, including social interactions, play, responses to their owner, and even being alone without their owner. Finally, high-quality and stable videos were essential for precise feature extraction and labeling.

B. *Large Language Model*

To analyze the videos and translate them into States and Actions for policy building within a Markov Decision Process framework, we employed a large language model (LLM). Specifically, we utilized the GPT-4o mini API, which supports image input for analysis. While individual frame analysis may not fully capture the dynamic nature, Madan et al. demonstrated in their work on video understanding tasks that image-based foundation models often tend to work better than video-based models in many cases [23]. Based on this approach, we utilized an LLM to analyze each frame image sent for processing.

Given that our dataset consists of videos, we extracted frames at consistent intervals and sent these images to GPT-4o mini API for analysis. Using an LLM to analyze every single frame from the videos to determine dogs’ states and actions is impractical. Instead, we sampled one frame every few seconds to ensure sufficient information for analysis while minimizing computational complexity. For instance, the videos from the GitHub repository we mentioned earlier have an average frame rate of 30 frames per second (FPS), and the YouTube video we utilized runs at 23 FPS. For these videos, we skipped approximately every 30 or 23 frames, respectively, and selected one frame per second for analysis.

In addition, prompt engineering played a crucial role in obtaining the desired outputs from GPT-4o mini API. The objective was to analyze each input frame and derive the

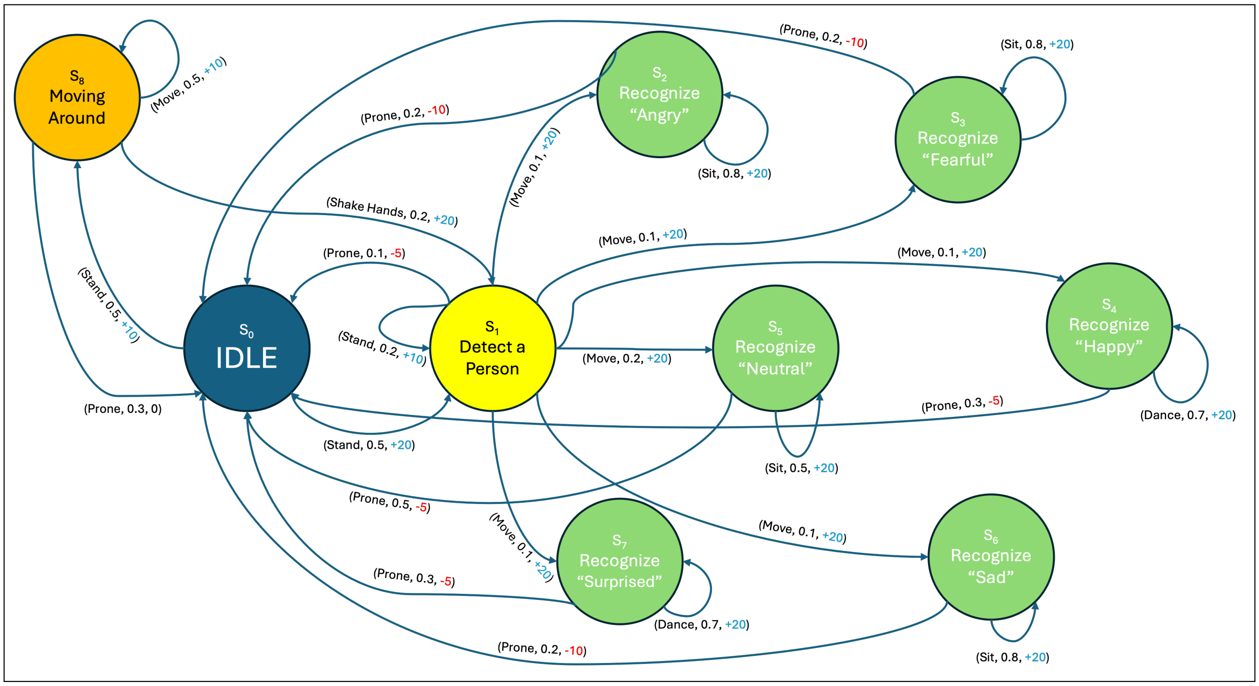

Fig. 1. State Transition Diagram Design of the Markov Decision Process (MDP) for the Animal Natural Behavior Learning Model

corresponding States and Actions of the dogs in the videos. To achieve accurate, reliable, and consistent results, we carefully designed and refined the prompts. Below are the prompts used in our program to analyze the dogs' States and Actions in the videos.

- “What is happening in this video? How is the dog behaving?”
- “This is for building a Markov Decision Process model for a robot dog.”
- “Please describe the state and action of the dog in the video.”
- “There is no such state or action ’none’.”
- “Please keep it concise and clear (in 1 word each) for each frame, such as ’State: [state] / Actions: [action]’.”
- “You don’t have to say anything like ’frame 1’, ’frame 2’, etc.”
- “You don’t have to say anything like ’Okay’, ’Sure’, etc.”

Even with the designed prompts, additional filtering was necessary to extract relevant terms for the States and Actions required for the robotic dog's natural behaviors. This involved ’normalizing’ the LLM's responses to ensure consistency and to align with the desired terms. For example, we standardized variations such as ’sitting,’ ’sitting down,’ or ’seated’ into a single term, ’Sit.’ However, due to the inherent flexibility of a large language model, it was not entirely possible to constrain its responses to consistently produce the exact terms we required.

C. Markov Decision Process

To develop the robotic dog's natural behavior policy, we employed a Markov Decision Process (MDP) framework using analyzed States and Actions derived from LLM-processed input frame images of living dogs. An MDP is a stochastic sequential decision process used for modeling “decision-making” problems [24].

In Figure 1, the majority of the states (represented by large circles) involve detecting and recognizing a person, while the remaining states include an IDLE state and a state where the robotic dog moves around randomly. Actions represent the possible behaviors of the robotic dog, such as ’Sit,’ ’Prone,’ ’Move,’ ’Shake Hands,’ ’Stand,’ or ’Dance.’ Each state transitions to a specific subsequent state, with the corresponding action determined by the context, as depicted in Figure 1.

Simultaneously, rewards were defined and assigned to observed transitions, as illustrated in Figure 1. The reward function was designed to prioritize behaviors that provide social and emotional support to the detected person. Favorable transitions, such as moving closer to a person, staying with them, or dancing with them—indicative of supportive behaviors—were assigned higher reward values. Conversely, neutral/undesirable transitions, such as detecting a person or recognizing their emotions but taking no action, received zero or even negative rewards.

The complete MDP model was assembled using the normalized states, actions, transition probabilities, and reward values, and was represented as a structured JSON object. Below are the key components of the MDP:

1) State (S): As briefly mentioned above, the State S represents the robot's current context, including posture, recognized pre-trained human emotions, environmental conditions, or the interaction history with humans.

2) Action (A): The Action A includes behaviors such as Move, Stand, Sit, Prone, Shake Hands, and other potential actions unique to the robotic dog.

3) Transition Probability (T): The Transition Probability function T utilizes observed interaction data, incorporating environmental uncertainties to model state transitions effectively.

4) Reward (R): The reward function R prioritizes behaviors that promote positive interactions, reinforcing desirable state transitions.

5) Policy: Policy = $\pi(a|S)$

The policy defines the probability of taking action a given the current state S, guiding the robot's decision-making process.

There are several examples based on the design above. For instance, if the robotic dog detects a person, moves closer, recognizes their emotion as happiness, and dances with the person, it receives positive rewards. Conversely, if the robotic dog detects a person who appears angry but merely lies prone without offering emotional support, it incurs negative rewards. Similarly, if the robotic dog detects a person but takes no action and remains prone, it also receives negative rewards. This flexible reward structure allows the MDP to effectively capture the priorities and objectives of the experimental design.

IV. Discussion

State/Action Results from LLM

The State/Action keywords for the robotic dog, extracted through LLM analysis, represent the core findings from the video data. To visualize the keywords, we generated word clouds illustrating key terms. These state and action-related keywords are essential for guiding the robot dog's movements, enabling it to exhibit more natural behaviors.

Figures 2, 3, 4, and 5 illustrate example results derived from the video datasets analyzed in this study. The word cloud in Figure 2 was generated from the video dataset provided by the public GitHub repository mentioned earlier [21], whereas the word clouds in Figures 3, 4, and 5 were created using data from YouTube videos [22].

Fig. 2. Word Cloud image of Dog's States and Actions from [21]'s videos.

Fig. 3. Word Cloud image of top 100 behaviors identified from a YouTube video [22].

Fig. 4. Word Cloud image of top 50 behaviors identified from the same source as Fig. 3.

Fig. 5. Word Cloud image of top 25 behaviors identified from the same source as Fig. 3.

Limitations & Future Works

This study utilizes the LLM (GPT-4o mini API), which generates responses that are not always identical or consistently formatted. As a result, additional filtering was also necessary to extract the desired results, even with prompting and normalization. While the outputs may vary slightly, they generally aligned with the given prompts. To ensure more consistent results in the desired format during program execution, further advancements in prompt design and processing methods are needed. As mentioned, frame images were extracted from the video every 1 second to maintain continuity and context and then sent to the LLM for analysis. However, as static representations, individual frames may not fully capture the dynamic nature the videos. Future studies could address this limitation by incorporating 3D pose analysis of dogs in videos and integrating this data with the LLM to achieve more accurate and comprehensive results.

Acknowledgment

This work was partially funded by the NSF Grant # 2229885 Institute for Trustworthy AI in Law and Society (TRAILS) as a seed project entitled "Trustworthy Human-Animal Robot Interaction (THARI) for Autistic Children."

References

[1] M. Botha, J. Hanlon, and G. L. Williams, "Does language matter? identity-first versus person-first language use in autism research: A response to vivanti," *Journal of Autism and Developmental Disorders*, vol. 53, no. 2, pp. 870–878, 2023.

[2] M. J. Maenner, "Prevalence and characteristics of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, united states, 2020," *MMWR. Surveillance Summaries*, vol. 72, 2023.

[3] F. Martin and J. Farnum, "Animal-assisted therapy for children with pervasive developmental disorders," *Western journal of nursing research*, vol. 24, no. 6, pp. 657–670, 2002.

[4] M. E. O'Haire, "Animal-assisted intervention for autism spectrum disorder: A systematic literature review," *Journal of autism and developmental disorders*, vol. 43, pp. 1606–1622, 2013.

[5] M. Bogliolo, F. Bogliolo, F. Operto, and M. Emanuele, "Nao: a promising tool for pediatric hospitals," in *2024 International Conference on Artificial Intelligence, Computer, Data Sciences and Applications (ACDSA)*. IEEE, 2024, pp. 1–6.

[6] A. Maignan, D. Xie, S. Ha, J. Kim, and C. H. Park, "Emotion-behavior interplay in human animal-robot interaction (hari)," in *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, 2024, pp. 727–731.

[7] Á. Miklósi and M. Gácsi, "On the utilization of social animals as a model for social robotics," *Frontiers in psychology*, vol. 3, p. 75, 2012.

[8] A. P. Association *et al.*, "Diagnostic and statistical manual of mental disorders," *Text revision*, 2000.

[9] A. Masi, M. M. DeMayo, N. Glozier, and A. J. Guastella, "An overview of autism spectrum disorder, heterogeneity and treatment options," *Neuroscience bulletin*, vol. 33, pp. 183–193, 2017.

[10] I. M. Tezcan, "Autism beyond defcits: Embracing neurodiversity in research."

[11] S. K. Kapp, *Autistic community and the neurodiversity movement: Stories from the frontline*. Springer Nature, 2020.

[12] E. Tomlinson and S. Newman, "Valuing writers from a neurodiversity perspective: Integrating new research on autism spectrum disorder into composition pedagogy," *Composition Studies*, vol. 45, no. 2, pp. 91–112, 2017.

[13] C. Wijker, R. Leontjevas, A. Spek, and M.-J. Enders-Slegers, "Effects of dog assisted therapy for adults with autism spectrum disorder: An exploratory randomized controlled trial," *Journal of autism and developmental disorders*, vol. 50, no. 6, pp. 2153–2163, 2020.

[14] U. G. Hunjan and K. J. Reddy, "Effect of animal-assisted therapy on depression, memory, attention, and emotion regulation," *Human Research in Rehabilitation*, vol. 13, no. 1, pp. 120–129, 2023.

[15] K. K. Hardy and R. N. Weston, "Canine-assisted therapy for children with autism spectrum disorder: A systematic review," *Review Journal of Autism and Developmental Disorders*, vol. 7, pp. 197–204, 2020.

[16] K. Silva, M. Lima, A. Santos-Magalhães, C. Fafiães, and L. de Sousa, "Living and robotic dogs as elicitors of social communication behavior and regulated emotional responding in individuals with autism and severe language delay: A preliminary comparative study," *Anthrozoös*, vol. 32, no. 1, pp. 23–33, 2019.

[17] K. Thodberg, L. U. Sørensen, P. B. Videbech, P. H. Poulsen, B. Houbak, V. Damgaard, I. Keseler, D. Edwards, and J. W. Christensen, "Behavioral responses of nursing home residents to visits from a person with a dog, a robot seal or atoy cat," *Anthrozoös*, vol. 29, no. 1, pp. 107–121, 2016.

[18] S. J. Chang, J. Lee, H. An, W.-H. Hong, and J. Y. Lee, "animal-assisted therapy as an intervention for older adults: A systematic review and meta-analysis to guide evidence-based practice," *Worldviews on Evidence-Based Nursing*, vol. 18, no. 1, pp. 60–67, 2021.

[19] G. Lakatos, "Dogs as behavior models for companion robots: How can human–dog interactions assist social robotics?" *IEEE Transactions on Cognitive and Developmental Systems*, vol. 9, no. 3, pp. 234–240, 2016.

[20] K. A. Riddoch, R. D. Hawkins, and E. S. Cross, "Exploring behaviours perceived as important for human—dog bonding and their translation to a robotic platform," *Plos one*, vol. 17, no. 9, p. e0274353, 2022.

[21] R. Loney, "dog-pose-estimation," GitHub, 2021. [Online]. Available: https://github.com/ryanloney/dog-pose-estimation

[22] Ring, "Pet Parents Tell Mama Dog via Their Ring Cam to Feed Her Puppies, Does She Listen? — RingTV," YouTube, 2022. [Online]. Available: https://youtu.be/HMAy9LznW7k?si=gPxyOQFxbyqVGE2p

[23] N. Madan, A. Moegelmose, R. Modi, Y. S. Rawat, and T. B. Moeslund, "Foundation models for video understanding: A survey," 2024. [Online]. Available: https://arxiv.org/abs/2405.03770

[24] M. L. Puterman, "Chapter 8 markov decision processes," in *Stochastic Models*, ser. Handbooks in Operations Research and Management Science. Elsevier, 1990, vol. 2, pp. 331–434. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0927050705801720