Towards Proactive Robot Learners that Ask for Help

Michelle Zhao

***Abstract*—Robots must learn how to behave in accordance with human desires and values. While today's robot learning algorithms increasingly enable people to teach robots via diverse forms of feedback (e.g., demonstration, language, etc.), they place the burden of responsibility on the human to perfectly understand what the robot doesn't know and provide the "right" data. Instead, I contend that robots should be proactive participants—they should bear some of the burden of knowing when they don't know and should ask for targeted help. I tackle this problem by extending foundational uncertainty quantification techniques to the HRI setting, enabling robots to rigorously "know when they don't know" even when relying on black-box policies. With these uncertainty quantification algorithms, I enable robots to ask for strategic help, like additional action labels during interactive imitation learning. My future work is focused on *how* robots should ask for help: developing algorithms for user-aware, fluent robot queries for online human feedback. I envision robots that explain what they are uncertain about in natural language, but also understand which feedback modalities (e.g. verbal clarification versus actions) might best resolve its uncertainty and are thus appropriate to ask for.**

***Index Terms*—learning from human feedback, uncertainty quantification, preference learning**

I. Introduction

One outstanding challenge in developing interactive autonomous systems is ensuring that robots operate in accordance with user intentions: assisting and interacting with people in the ways users prefer [1]–[3]. Ultimately, the evaluator of any interactive robot is an end-user: a tidy-bot [4] should rearrange your belongings the way you like and an AI assistant should sound like you when helping your write e-mails [5], [6]. However, today's robots don't act according to human preferences "out of the box" [7]–[9]. Thus, human oversight remains crucial during a robot's life-cycle. For example, robots are aligned with end-user by explicitly getting their feedback, whether it be preferences [10]–[12], demonstrations [13], corrections [14], language [15], or implicitly communicative behaviors [16]–[18]. Even after alignment, robots are supervised by operators, such as safety drivers monitor autonomous vehicles [19]–[21] and everyday people who constantly check if LLM outputs are reliable enough to use in their emails [22]–[24]. In other words, humans bear the burden of deciphering what robot's know, what they don't know, and preventing their failures with targeted data [25], [26].

Instead, I believe that robots should know when they don't know and ask for targeted help, shouldering some of the burden of preempting potential errors and helping the user provide actionable feedback. To accomplish this, advances in uncertainty quantification is a crucial first step. Robots must first know **when they are uncertain**. While Bayesian approaches [27]–[30] afford uncertainty quantification via posterior entropy, leveraged as well in my prior work

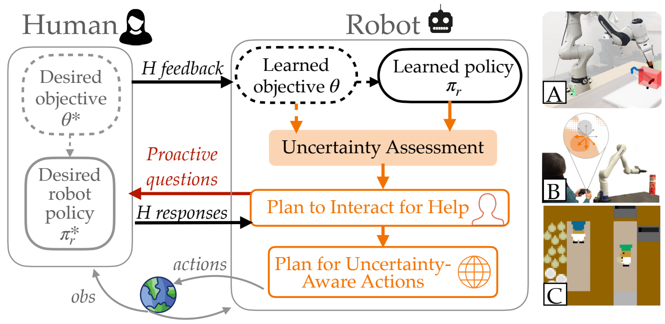

Fig. 1. The robot seeks to align its learned policy, π_r, which may include modeling of objective θ. My research aims to facilitate proactive questions (red) between robot and human, enabling the robot to engage in a *collaborative* alignment process, situated in real (A,B) and simulated domains (C).

[16]–[18], [31], [32], an open challenge is how to quantify robot decision uncertainty in the presence of the end-to-end "black box" policies underlying imitation and reinforcement learning. To tackle this, I draw from statistical uncertainty quantification techniques, and contributed a model-agnostic method for robots to measure policy-level uncertainty in their predicted actions [33], *even* when human feedback is observed intermittently [34], providing the robot a self-assessment of uncertainty, calibrated to human feedback received online. Equipped with uncertainty estimates, the second step is for robots to use their uncertainty to guide **when to ask for help**. I developed an interactive imitation learning approach wherein the robot uses its calibrated uncertainty as a reliable measure of deployment-time uncertainty to actively query for more expert feedback [34], allowing the robot to become proactive in seeking intervention via expert control before erroneous actions are performed and quickly adapt its behavior.

My remaining work will focus on the question of **how to interact for help**: designing user-friendly methods for robots to communicate uncertainty [32], [35] and create understandable, multimodal queries for feedback. The ultimate goal of my thesis is to enable robots to be proactive participants in the alignment process: assessing uncertainty and seeking continual feedback even during online execution, while also being cognizant of the user's capacity for feedback and interaction preferences. I envision robots that quantify and communicate their uncertainty with in-the-loop users, but also ask questions via different feedback modalities (verbal clarification versus action control) that suit different sources of uncertainty.

II. Quantifying Uncertainty and Asking for Help

My research is focused on how robot learners can ask for help when they are uncertain. This is motivated by my early works in which I observed how uncertainty plays a key role in collaborative robot decision-making. My early work showed robots can approach collaboration (in a simulated collaborative cooking task Fig 1C) through uncertainty-aware decisions [16] via a Mixture-of-Experts policy, weighting actions of multiple reward-specific policies by the likelihood of its reward, where Bayesian posterior entropy affords a measure of uncertainty over rewards. Next, I [17] showed planning ahead using an optimistic information gain objective helps the robot refrain from hastily performing actions that restrict future choices of the human, allowing the robot to more effectively resolve uncertainty in its beliefs. However, these methods were grounded in low-dimensional robot policies, and highlights the need for effective high-dimensional uncertainty quantification techniques to guide interactive decisions.

Extending Distribution-free Uncertainty Quantification to HRI. To ensure that its behaviors are aligned with user desires, the robot's uncertainty lies in whether its policy is producing desired behaviors [36]. Although methods like ensembles [37], [38] quantify uncertainty stemming from the training data, the robot's uncertainty estimate can become quickly uncalibrated as it encounters distribution shift or changing user intentions during execution. Instead, I contend that the human feedback requested and received *during* deployment time is a valuable uncertainty quantification (UQ) signal that should be leveraged to *update* the robot's uncertainty estimate *online*. If properly accounted for, the updated uncertainty estimate will influence when the robot asks for more help, enabling it to targetedly query the human to improve policy performance.

I introduce two methods for adaptive uncertainty quantification over robot policies: given intermittent human feedback (IQT [34]) and constant human feedback (ACQR [33]). My approach is grounded in online conformal prediction [39] which is a distribution-free, model-agnostic way to represent uncertainty (with statistical guarantees) via prediction intervals constructed on the output of the robot's policy. Based on human feedback via action labels, IQT and ACQR grow and shrink uncertainty intervals over time as the user's intention is poorly- or well-predicted by the robot's policy (Fig 1 A).

ACQR provides a critical validation step for safe deployment of learned policies for assistive teleoperation [40], [41], where data-driven methods [42] learn a mapping from low-DoF human inputs to high-DoF robot actions. Given the learned assistive policy, and a small dataset of inputs and desired actions from target end-users, ACQR detects critical high-uncertainty states and inputs for both in- and out-of-distribution end-user behavior (Fig 1 B, high/low uncertainty input regions in orange/gray). IQT lifts the assumption in prior online conformal approaches, including ACQR, of constant feedback, which would be difficult for real human users to provide. In doing so, it enables rigorous *deployment-time* uncertainty quantification based on sparse user feedback. IQT's statistical coverage guarantees ensure that uncertainty is *calibrated* to distribution shifts in environment or human intentions. IQT is model-agnostic, enabling UQ on both structured (e.g. Bayesian) approaches and black-box (e.g. LLM- and diffusion-based) policies. Critically, IQT enables robots to measure uncertainty in the predicted actions of learned policies, allowing them to be proactive in seeking intervention.

Robots that Ask for Help during Interactive Imitation Learning. Next, I turned my attention to how the robot can use its uncertainty estimate to guide interaction with the user: asking for human feedback in the form of expert control. I develop ConformalDAgger [34], a new interactive IL approach wherein the robot uses prediction intervals calibrated by IQT as a reliable measure of deployment-time uncertainty to actively query for more expert feedback. ConformalDAgger uniquely combines human-gated and robot-gated feedback: users can anytime independently provide feedback to the robot, and the robot can actively request human control when uncertain. ConformalDAgger automatically increases uncertainty online when the expert shifts, resulting in more expert label queries compared to EnsembleDAgger [43] and allowing our approach to rapidly learn a policy aligned with the expert's intentions.

III. Future Work: How to Interact for Help

My current research established the algorithmic foundations for robots to know when they don't know, even when they use black-box robot policies during human interaction. However, one thing my early research studies [17] showed was that participants still lacked an understanding of the robot's uncertainty, and the robot's learned behaviors still often misaligned with users' expectation. This motivates my future work to study how robots should communicate and interact for help: (1) developing algorithms for proactive dialogue when the robot is uncertain and formalizing multimodal feedback, (2) personalizing queries to task constraints and user preferences.

Proactive Dialogue. During the interaction, the robot isn't limited to seeking alignment only through physical actions, but also through interactive, in-the-moment queries to the user during ongoing task execution [44]. I will develop an information-theoretic formalization of different types of *online verbal and physical* queries, augmenting the robot's means for asking for help (e.g. confirmatory queries that invite but do not require responses: "I'll clear the dishes") and *explaining why* it's uncertain. I take a first step towards proactive dialogue via multimodal explanations of multiagent decision-making [32].

Asking User-Cognizant Queries. Second, I want to ensure that robots pose questions that are consequential [45], [46] with respect to the task and in ways that are preferable to or easy to answer for the user. Users may at times require the robot to perform learned tasks independently and at others be available for teaching. How can the robot targetedly ask about task-critical uncertainty through user-preferred query types? I plan to augment the robot's task representation to account for user interaction objectives and evaluate via user studies the impact of user-aware learners that ask for feedback accounting for task consequentiality and user interaction preferences.

I ultimately want robots to view the process of learning behaviors aligned with human objectives as collaborative, empowering in-the-loop users to better understand the robot's capabilities, provide more effective, intuitive feedback, and confidently teach robot learners to align with user objectives.

References

[1] D. S. Brown, J. Schneider, A. Dragan, and S. Niekum, "Value alignment verification," in *International Conference on Machine Learning*. PMLR, 2021, pp. 1105–1115.

[2] I. Gabriel and V. Ghazavi, "The challenge of value alignment: From fairer algorithms to ai safety," *arXiv preprint arXiv:2101.06060*, 2021.

[3] D. Hadfield-Menell, S. J. Russell, P. Abbeel, and A. Dragan, "Cooperative inverse reinforcement learning," *Advances in neural information processing systems*, vol. 29, 2016.

[4] J. Wu, R. Antonova, A. Kan, M. Lepert, A. Zeng, S. Song, J. Bohg, S. Rusinkiewicz, and T. Funkhouser, "Tidybot: Personalized robot assistance with large language models," *Autonomous Robots*, vol. 47, no. 8, pp. 1087–1102, 2023.

[5] K. Greshake, S. Abdelnabi, S. Mishra, C. Endres, T. Holz, and M. Fritz, "Not what you've signed up for: Compromising real-world llm-integrated applications with indirect prompt injection," in *Proceedings of the 16th ACM Workshop on Artificial Intelligence and Security*, 2023, pp. 79–90.

[6] A. Bastola, H. Wang, J. Hembree, P. Yadav, N. McNeese, and A. Razi, "Llm-based smart reply (lsr): Enhancing collaborative performance with chatgpt-mediated smart reply system (acm)(draft) llm-based smart reply (lsr): Enhancing collaborative performance with chatgpt-mediated smart reply system," *arXiv preprint arXiv:2306.11980*, 2023.

[7] C. Spiess, D. Gros, K. S. Pai, M. Pradel, M. R. I. Rabin, A. Alipour, S. Jha, P. Devanbu, and T. Ahmed, "Calibration and correctness of language models for code," *arXiv preprint arXiv:2402.02047*, 2024.

[8] T. Zhao, M. Wei, J. S. Preston, and H. Poon, "Automatic calibration and error correction for large language models via pareto optimal self-supervision," *arXiv preprint arXiv:2306.16564*, 2023.

[9] A. Mitrevski, S. Thoduka, A. O. Sáinz, M. Schöbel, P. Nagel, P. G. Plöger, and E. Prassler, "Deploying robots in everyday environments: Towards dependable and practical robotic systems," *arXiv preprint arXiv:2206.12719*, 2022.

[10] D. Sadigh, A. D. Dragan, S. Sastry, and S. A. Seshia, *Active preference-based learning of reward functions*, 2017.

[11] E. Biyik and D. Sadigh, "Batch active preference-based learning of reward functions," *CoRR*, vol. abs/1810.04303, 2018. [Online]. Available: http://arxiv.org/abs/1810.04303

[12] T. Fitzgerald, P. Koppol, P. Callaghan, R. Q. J. H. Wong, R. Simmons, O. Kroemer, and H. Admoni, "Inquire: Interactive querying for user-aware informative reasoning," in *6th Annual Conference on Robot Learning*, 2022.

[13] S. A. Mehta and D. P. Losey, "Unified learning from demonstrations, corrections, and preferences during physical human-robot interaction," *arXiv preprint arXiv:2207.03395*, 2022.

[14] A. Bajcsy, D. P. Losey, M. K. O'Malley, and A. D. Dragan, "Learning from physical human corrections, one feature at a time," in *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, 2018, pp. 141–149.

[15] Z. Yang, M. Jun, J. Tien, S. Russell, A. Dragan, and E. Biyik, "Trajectory improvement and reward learning from comparative language feedback," in *8th Annual Conference on Robot Learning*.

[16] M. Zhao, R. Simmons, and H. Admoni, "Coordination with humans via strategy matching," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 9116–9123.

[17] ——, "Learning human contribution preferences in collaborative human-robot tasks," in *Conference on Robot Learning*. PMLR, 2023, pp. 3597–3618.

[18] ——, "The role of adaptation in collective human-ai teaming," *Topics in Cognitive Science*, 2022.

[19] M. L. Cummings, "What self-driving cars tell us about ai risks," *IEEE Spectrum*. [Online]. Available: https://spectrum.ieee.org/self-driving-cars-2662494269

[20] W. Osten, C. Bett, and G. Situ, "The challenge of making self-driving cars: may ai help to overcome the risks, or should we focus on reliable sensor technologies?" in *Interferometry and Structured Light 2024*, vol. 13135. SPIE, 2024, pp. 8–21.

[21] A. L. Stein, "Assuming the risks of artificial intelligence," *BUL Rev.*, vol. 102, p. 979, 2022.

[22] Y. Chang, X. Wang, J. Wang, Y. Wu, L. Yang, K. Zhu, H. Chen, X. Yi, C. Wang, Y. Wang *et al.*, "A survey on evaluation of large language models," *ACM Transactions on Intelligent Systems and Technology*, vol. 15, no. 3, pp. 1–45, 2024.

[23] J. Chervenak, H. Lieman, M. Blanco-Breindel, and S. Jindal, "The promise and peril of using a large language model to obtain clinical information: Chatgpt performs strongly as a fertility counseling tool with limitations," *Fertility and sterility*, vol. 120, no. 3, pp. 575–583, 2023.

[24] A. Hamidi and K. Roberts, "Evaluation of ai chatbots for patient-specific ehr questions," *arXiv preprint arXiv:2306.02549*, 2023.

[25] E. Hedlund, M. Johnson, and M. Gombolay, "The effects of a robot's performance on human teachers for learning from demonstration tasks," in *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, 2021, pp. 207–215.

[26] A. L. Thomaz and C. Breazeal, "Teachable robots: Understanding human teaching behavior to build more effective robot learners," *Artificial Intelligence*, vol. 172, no. 6-7, pp. 716–737, 2008.

[27] Z. Chen *et al.*, "Bayesian filtering: From kalman filters to particle filters, and beyond," *Statistics*, vol. 182, no. 1, pp. 1–69, 2003.

[28] E. Goan and C. Fookes, "Bayesian neural networks: An introduction and survey," *Case Studies in Applied Bayesian Data Science: CIRM Jean-Morlet Chair, Fall 2018*, pp. 45–87, 2020.

[29] J. Lampinen and A. Vehtari, "Bayesian approach for neural networks—review and case studies," *Neural networks*, vol. 14, no. 3, pp. 257–274, 2001.

[30] I. Kononenko, "Bayesian neural networks," *Biological Cybernetics*, vol. 61, no. 5, pp. 361–370, 1989.

[31] M. Tecson, D. Chen, M. Zhao, R. Simmons, and Z. Erikson, "Leveraging large language models for preference-based sequence prediction," in *Under Review, International Conference on Agents and Artificial Intelligence (ICAART)*. ICAART, 2025.

[32] R. Pandya, M. Zhao, C. Liu, R. Simmons, and H. Admoni, "Multi-agent strategy explanations for human-robot collaboration," in *IEEE International Conference on Robotics and Automation*. ICRA, 2024.

[33] M. Zhao, R. Simmons, H. Admoni, and A. Bajcsy, "Conformalized teleoperation: Confidently mapping human inputs to high-dimensional robot action," in *Robotics: Science and Systems (RSS)*. RSS, 2024.

[34] M. Zhao, R. Simmons, H. Admoni, A. Ramdas, and A. Bajcsy, "Conformalized interactive imitation learning: Handling expert shift and intermittent feedback," in *Under review, International Conference on Learning Representations (ICLR)*. ICLR, 2025.

[35] S. Habibian, A. A. Valdivia, L. H. Blumenschein, and D. P. Losey, "A review of communicating robot learning during human-robot interaction," *arXiv preprint arXiv:2312.00948*, 2023.

[36] A. Jauffret, C. Grand, N. Cuperlier, P. Gaussier, and P. Tarroux, "How can a robot evaluate its own behavior? a neural model for self-assessment," in *The 2013 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2013, pp. 1–8.

[37] J. Mendes-Moreira, C. Soares, A. M. Jorge, and J. F. D. Sousa, "Ensemble approaches for regression: A survey," *Acm computing surveys (csur)*, vol. 45, no. 1, pp. 1–40, 2012.

[38] B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and scalable predictive uncertainty estimation using deep ensembles," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[39] I. Gibbs and E. Candes, "Adaptive conformal inference under distribution shift," *Advances in Neural Information Processing Systems*, vol. 34, pp. 1660–1672, 2021.

[40] S. W. Brose, D. J. Weber, B. A. Salatin, G. G. Grindle, H. Wang, J. J. Vazquez, and R. A. Cooper, "The role of assistive robotics in the lives of persons with disability," *American Journal of Physical Medicine & Rehabilitation*, vol. 89, no. 6, pp. 509–521, 2010.

[41] D. P. Miller, "Assistive robotics: an overview," *Assistive Technology and Artificial Intelligence: Applications in Robotics, User Interfaces and Natural Language Processing*, pp. 126–136, 2006.

[42] A. Jonnavittula, S. A. Mehta, and D. P. Losey, "Learning to share autonomy from repeated human-robot interaction," *arXiv preprint arXiv:2205.09795*, 2022.

[43] K. Menda, K. Driggs-Campbell, and M. J. Kochenderfer, "Ensembledagger: A bayesian approach to safe imitation learning," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 5041–5048.

[44] C. Lynch and P. Sermanet, "Language conditioned imitation learning over unstructured data," *arXiv preprint arXiv:2005.07648*, 2020.

[45] S. H. Huang, K. Bhatia, P. Abbeel, and A. D. Dragan, "Establishing appropriate trust via critical states," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 3929–3936.

[46] S. Sreedharan, S. Srivastava, and S. Kambhampati, "Tldr: Policy summarization for factored ssp problems using temporal abstractions," in *Proceedings of the International Conference on Automated Planning and Scheduling*, vol. 30, 2020, pp. 272–280.