Causal-HRI: Causal Learning for Human-Robot Interaction

Jiaee Cheong
jc2208@cam.ac.uk
University of Cambridge, UK,
The Alan Turing Institute, UK

Nikhil Churamani
nikhil.churamani@cl.cam.ac.uk
University of Cambridge, UK

Luke Guerdan
lguerdan@cs.cmu.edu
Carnegie Mellon University, USA

Tabitha Edith Lee
tabithalee@cmu.edu
Carnegie Mellon University, USA

Zhao Han
zhaohan@usf.edu
University of South Florida, USA

Hatice Gunes
hatice.gunes@cl.cam.ac.uk
University of Cambridge, UK

ABSTRACT

Real-world Human-Robot Interaction (HRI) requires robots to adeptly perceive and understand the dynamic human-centred environments in which they operate. Recent decades have seen remarkable advancements that have endowed robots with exceptional perception capabilities. The first workshop on "Causal-HRI: Causal Learning for Human-Robot Interaction" aims to bring together research perspectives from *Causal Discovery and Inference* and *Causal Learning*, in general, to real-world HRI applications. The objective of this workshop is to explore strategies that will not only embed robots with capabilities to *discover cause-and-effect relationships* from observations, allowing them to generalise to unseen interaction settings, but also to enable users to *understand robot behaviours*, moving beyond the 'black-box' models used by these robots. This workshop aims to facilitate an exchange of views through invited keynote presentations, contributed talks, group discussions and poster sessions, encouraging collaborations across diverse scientific communities. The theme of HRI 2024, "HRI in the real world," will inform the overarching theme of this workshop, encouraging discussions on HRI theories, methods, designs and studies focused on leveraging Causal Learning for enhancing real-world HRI.

CCS CONCEPTS

• **Computing methodologies** → *Causal reasoning and diagnostics*; *Intelligent agents*; *Cognitive robotics*; *Machine learning*; • **Human-centered computing**;

KEYWORDS

Causal Discovery, Causal Inference, Causal Learning, Cognitive Robotics, Intelligent Agents, Human-Robot Interaction, Robotics

ACM Reference Format:
Jiaee Cheong, Nikhil Churamani, Luke Guerdan, Tabitha Edith Lee, Zhao Han, and Hatice Gunes. 2024. Causal-HRI: Causal Learning for Human-Robot Interaction. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction (HRI '24 Companion), March 11–14, 2024, Boulder, CO, USA.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3610978.3638157

HRI '24 Companion, March 11–14, 2024, Boulder, CO, USA

ACM ISBN 979-8-4007-0323-2/24/03.
https://doi.org/10.1145/3610978.3638157

1 INTRODUCTION

The successful deployment of robots that collaborate with humans in real-world environments must be addressed by machine intelligence that understands not only environment objects and features, but also human interaction. From coordination to cooperation, robots that accommodate human preferences and needs to masterfully operate in the real world require learning and reasoning capabilities that exceed the intelligence that can currently be endowed with state-of-the-art methods based on correlation. Yet, causal reasoning and learning, hallmarks of human intelligence [9–11, 17], promise a path forward to the robot cognition required for the vast social benefits that emerge from successful robot deployments in homes, hospitals, and other public places.

Indeed, early attempts at leveraging causality to improve robotic capabilities have proven promising [14, 15]. Causal inference methods have been shown to improve the robustness of robot policies learned from human demonstrations [19, 20], explain robot failures [5], learn robot intent communication [7, 12] and even discover the world dynamics from human interactions [2, 6].

As the integration of causality within machine learning (ML) is still nascent [18], we believe the field of Human-Robot Interaction (HRI) stands to gain by examining and integrating new methods in causal learning. To this end, the "Causal-HRI: Causal Learning for Human-Robot Interaction" Workshop[1] seeks to bring together a multidisciplinary team of researchers to highlight and discuss novel methods in *causal* HRI and identify future directions in this exciting and promising new field.

2 BACKGROUND

Recent research developments have resulted in robotic systems becoming more ubiquitous in human life. These robots are now increasingly taking on roles that involve complex HRI dynamics (for example, assistants and tutors). However, much of this progress is grounded in pattern recognition and statistical correlation-based ML, neglecting the intrinsic structures and interdependencies between variables in observational data and the underlying causal relationships that govern the emergence of these dependencies. Causality focuses on unraveling such *causal structures and relationships* inherent in the data. Many challenges within ML and HRI, including *generalisation* and *bias* issues, can be attributed to this ignorance of cause-and-effect relationships between data variables.

Tools within causal reasoning such as the causal hierarchy (Association, Intervention and Counterfactuals) [16], causal relationship

[1]The website of the Causal-HRI Workshop: https://causal-hri.github.io/

learning [3] and causal discovery methods [8] have posited that the challenges within both ML and HRI can be addressed by such tools [4, 21]. The recent workshop on "Causality for Robotics: Answering the Question of Why"[2] organised at IROS 2023 drew a range of submissions, highlighting how methods grounded in causality can address challenges in robotics research [1, 13]. This workshop attempts to extend these findings by focusing on HRI research.

3 WORKSHOP OVERVIEW

Causal-HRI is a half-day, hybrid workshop focused on exploring Causal Discovery and Inference and Causal Learning for real-world Human-Robot Interaction. The proposed workshop includes:

- **Keynote Talks:** Prof. Holly Yanco (University of Massachusetts Lowell), Prof. Alison Gopnik (University of California at Berkeley), and Prof. Karinne Ramirez-Amaro (Chalmers University of Technology, Sweden) will present their insights from Causal Discovery and Inference, Causal Representation Learning, Robotics, and Human-Robot Interaction. Keynote presentations will last 30 minutes (20-minutes talk with a 10-minute Q&A).
- **Contributed Talks:** The authors of accepted research *papers* will present their work as an 8-minute oral presentation, followed by 2-minute Q&A.
- **Poster Session:** The authors of the accepted *poster* contributions will present their position extended abstracts as *lightning talks* during the poster session.
- **Group Discussion:** Following the presentations, the audience will be split into smaller groups to facilitate discussions on the key themes of the workshop. Insights from the discussions will be collated and shared on the workshop website.

3.1 Target Audience and Advertisement for Participation

We invite authors to submit their contributions as 3-4 page (plus additional pages for references and appendices) *papers*, highlighting their experimental results, technical reports, and case studies focused on Causal Learning for Human-Robot Interaction. In particular, we encourage submissions addressing the theme of HRI 2024: "HRI in the real-world." All submissions will be peer-reviewed for their novelty, relevance, contribution to the field, and technical soundness. We also invite researchers to submit position articles as 1-2 page extended abstracts (*posters*). These accepted *poster* submissions will be presented as *lightning talks* during the dedicated poster session at the workshop.

The workshop is advertised to members of the Causality and Causal Learning community using the respective mailing lists and Slack channels such as ContinualCausality and Alan Turing Causal Inference Interest group, amongst others. Additionally, the workshop is advertised to various robotics and Human-Robot Interaction communities such as robotics-worldwide, HRI-Announcements, CHI-Announcements, and other dedicated working groups. The workshop is also announced using a dedicated website[3] and Slack Workspace[4] to form a community of researchers working on Causal Learning for Human-Robot Interaction. The workshop will also be advertised on social media channels such as Twitter (X) and LinkedIN. The workshop is expected to garner the attention of around 45-50 attendees from the Causal Learning, Causal Discovery and Inference, ML, and HRI communities.

3.2 Plans for Documenting the Workshop

Accepted papers will be published on the workshop website. Based on authors' consent and preferences, the proceedings of the workshop will be compiled as a single submission or as an indexed compendium of individual papers and made available on arXiv. As Causality for HRI is a relatively new topic, we would also gather the insights learnt during the workshop, in the form of an article submission for a robotics or HRI conference or journal to create a stepping stone for the community to start exploring these ideas.

3.3 List of Topics

Topics of interest include, but are not limited to:

- Causal inference and representation learning
- Scene understanding with causal inference
- Causal learning for human behaviour understanding
- Causal learning for skill-discovery for robots
- Causal discovery of latent graphs for robotic behaviour learning
- Causal learning for state/action-space inferences
- Counterfactual reasoning for robotics
- Generalised representation learning for HRI
- Explanations for robot behaviours
- Explainable human-robot interaction
- Applications for/of causal HRI
- Research datasets, software, open-source tools, hardware analysis, system benchmarks in/for causal HRI.

3.4 Statement of Inclusion, Diversity and Equity

The workshop will devote particular care to include traditionally underrepresented, historically marginalised and economically underprivileged attendees. To invite a diverse audience, the workshop is also advertised in affinity group organisations such as Black in AI, Indigenous in AI, LatinX in AI, Queer in AI, Women in ML, amongst others. To further improve global accessibility, the workshop is organised in a hybrid fashion and the recordings of the workshop will be made available publicly, post-conference.

4 ORGANISERS

Jiaee Cheong (University of Cambridge, UK) is a PhD student at the Affective Intelligence and Robotics (AFAR) Lab, University of Cambridge. Her research interests lie at the intersection of causality, fairness, machine learning, affective computing, and HRI.

Nikhil Churamani (University of Cambridge, UK) is a Postdoctoral Researcher at the AFAR Lab of the Department of Computer Science and Technology, University of Cambridge. His PhD research at the University of Cambridge focused on Continual Learning for Affective Robotics, funded by EPSRC, UKRI. His current research investigates Continual Learning of Affect for social robots, focused on affect-driven learning for Human-Robot Interaction as well as Federated Continual Learning of socially appropriate robot behaviors in human-centered environments. He has published in several

[2] https://sites.google.com/view/iros23-causal-robots
[3] https://causal-hri.github.io
[4] https://causal-hri.slack.com

top journals and conferences such as PMLR, IEEE Transactions of Affective Computing, Frontiers in Robotics & AI, ACM/IEEE HRI, IEEE FG, IEEE RO-MAN, IEEE IROS, amongst others.

Luke Guerdan (Carnegie Mellon University, USA) is a Ph.D. student in the Human-Computer Interaction Institute at Carnegie Mellon University. Luke conducts research at the intersection of causal inference, human-computer interaction, and machine learning, with an emphasis on evaluating the reliability and safety of AI systems. Luke completed his MPhil in the Affective Intellegence and Robotics Lab at the University of Cambridge under the supervision of Prof. Gunes, where his thesis was titled Federated Continual Learning for Human-Robot Interaction.

Tabitha Edith Lee (Carnegie Mellon University, USA) is a Ph.D. candidate in Robotics at Carnegie Mellon University's Robotics Institute. She is a member of the Intelligent Autonomous Manipulation lab and is advised by Prof. Oliver Kroemer. Her thesis research investigates causal robot learning for manipulation: the interplay between robot perception and control through the lens of causality to learn and leverage the causal structure of manipulation tasks. She was the lead organizer for the "Causality for Robotics: Answering the Question of Why" workshop at IROS 2023. Her research in structural sim-to-real transfer has been recognized by an Honorable Mention selection for the NCWIT Collegiate Award. She is also a Siebel Scholar in Computer Science.

Zhao Han (University of South Florida, USA) is an Assistant Professor of Computer Science and Engineering at the University of South Florida, and leads the Reality, Autonomy, and Robot Experience (RARE) Lab. His research lies broadly in HRI, robotics, AI, and augmented reality (AR). He received the best long-paper award at INLG 2022, the best late-breaking report third prize at HRI 2022, and a best late-breaking report nominee at HRI 2023. Co-editing several special journal issues, he also co-organized multiple workshops at HRI and ACII, and chaired paper sessions at IROS and AI-HRI.

Hatice Gunes (University of Cambridge, UK) is an internationally recognized scholar and a Full Professor of Affective Intelligence and Robotics. She is a former President of the Association for the Advancement of Affective Computing and was a Faculty Fellow of the Alan Turing Institute – UK's national centre for data science and artificial intelligence. She obtained her PhD in computer science from the University of Technology Sydney (UTS) in Australia as an awardee of the Australian Government International Postgraduate Research Scholarship. Now directing the AFAR Lab at the University of Cambridge's Department of Computer Science and Technology, Prof Gunes spearheads research on multimodal, social, and affective intelligence for AI systems, particularly embodied agents and robots, by cross-fertilizing research in the fields of Machine Learning, Affective Computing and Social Signal Processing and Human Nonverbal Behaviour Understanding.

ACKNOWLEDGEMENTS

Funding: J. Cheong is supported by the Alan Turing Institute Doctoral Studentship and the Leverhulme Trust. N. Churamani is supported by Google under the GIG Funding Scheme. L. Guerdan is supported by the National Science Foundation Graduate Research Fellowship Program (Award No. DGE1745016). H. Gunes is supported in part by the EPSRC/UKRI under grant ref. EP/R030782/1 (ARoEQ) and in part by Google under the GIG Funding Scheme.

REFERENCES

[1] Ricardo Cannizzaro, Jonathan Routley, and Lars Kunze. 2023. Towards a Causal Probabilistic Framework for Prediction, Action-Selection & Explanations for Robot Block-Stacking Tasks. *arXiv preprint arXiv:2308.06203* (2023).
[2] Luca Castri, Sariah Mghames, and Nicola Bellotto. 2023. From Continual Learning to Causal Discovery in Robotics. In *AAAI Bridge Program on Continual Causality*. PMLR, 85–91.
[3] Lu Cheng, Ruocheng Guo, Raha Moraffah, Paras Sheth, Kasim Selcuk Candan, and Huan Liu. 2022. Evaluation Methods and Measures for Causal Learning Algorithms. *IEEE Transactions on Artificial Intelligence* (2022).
[4] Nikhil Churamani, Jiaee Cheong, Sinan Kalkan, and Hatice Gunes. 2023. Towards Causal Replay for Knowledge Rehearsal in Continual Learning. In *AAAI Bridge Program on Continual Causality*. PMLR, 63–70.
[5] Maximilian Diehl and Karinne Ramirez-Amaro. 2022. Why Did I Fail? A Causal-Based Method to Find Explanations for Robot Failures. *IEEE Robotics and Automation Letters* 7, 4 (2022), 8925–8932.
[6] Filip Edström, Thomas Hellström, and Xavier De Luna. 2023. Robot Causal Discovery Aided by Human Interaction. In *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 1731–1736.
[7] Xiaofeng Gao, Ran Gong, Yizhou Zhao, Shu Wang, Tianmin Shu, and Song-Chun Zhu. 2020. Joint Mind Modeling for Explanation Generation in Complex Human-Robot Collaborative Tasks. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 1119–1126.
[8] Clark Glymour, Kun Zhang, and Peter Spirtes. 2019. Review of Causal Discovery Methods Based on Graphical Models. *Frontiers in Genetics* 10 (2019), 524.
[9] Alison Gopnik, Clark Glymour, David M Sobel, Laura E Schulz, Tamar Kushnir, and David Danks. 2004. A Theory of Causal Learning in Children: Causal Maps and Bayes Nets. *Psychological Review* 111, 1 (2004), 3.
[10] Thomas L Griffiths and Joshua B Tenenbaum. 2005. Structure and Strength in Causal Induction. *Cognitive Psychology* 51, 4 (2005), 334–384.
[11] Thomas L Griffiths and Joshua B Tenenbaum. 2009. Theory-Based Causal Induction. *Psychological Review* 116, 4 (2009), 661.
[12] Zhao Han, Boyoung Kim, Holly A Yanco, and Tom Williams. 2022. Causal Robot Communication Inspired by Observational Learning Insights. *arXiv preprint arXiv:2203.09114* (2022).
[13] Peide Huang, Xilun Zhang, Ziang Cao, Shiqi Liu, Mengdi Xu, Wenhao Ding, Jonathan Francis, Bingqing Chen, and Ding Zhao. 2023. What Went Wrong? Closing the Sim-to-Real Gap via Differentiable Causal Discovery. In *Conference on Robot Learning*. PMLR, 734–760.
[14] Tabitha Edith Lee, Shivam Vats, Siddharth Girdhar, and Oliver Kroemer. 2023. SCALE: Causal Learning and Discovery of Robot Manipulation Skills using Simulation. In *Conference on Robot Learning*. PMLR, 2229–2256.
[15] Tabitha Edith Lee, Jialiang Alan Zhao, Amrita S. Sawhney, Siddharth Girdhar, and Oliver Kroemer. 2021. Causal Reasoning in Simulation for Structure and Transfer Learning of Robot Manipulation Policies. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*. 4776–4782.
[16] Judea Pearl. 2009. *Causality*. Cambridge University Press.
[17] Derek C Penn and Daniel J Povinelli. 2007. Causal Cognition in Human and Nonhuman Animals: A Comparative, Critical Review. *Annual Review of Psychology* 58 (2007), 97–118.
[18] Bernhard Schölkopf, Francesco Locatello, Stefan Bauer, Nan Rosemary Ke, Nal Kalchbrenner, Anirudh Goyal, and Yoshua Bengio. 2021. Toward Causal Representation Learning. *Proc. IEEE* 109, 5 (2021), 612–634.
[19] Gokul Swamy, Sanjiban Choudhury, Drew Bagnell, and Steven Wu. 2022. Causal Imitation Learning under Temporally Correlated Noise. In *International Conference on Machine Learning*. PMLR, 20877–20890.
[20] Gokul Swamy, Sanjiban Choudhury, J Bagnell, and Steven Z Wu. 2022. Sequence Model Imitation Learning with Unobserved Contexts. *Advances in Neural Information Processing Systems* 35 (2022), 17665–17676.
[21] Kun Zhang, Bernhard Schölkopf, Peter Spirtes, and Clark Glymour. 2018. Learning Causality and Causality-Related Learning: Some Recent Progress. *National Science Review* 5, 1 (2018), 26–29.