

REFERENCES

- [1] DE Asher, SL Barton, E Zaroukian, and NR Waytowich. 2019. Effect of cooperative team size on coordination in adaptive multi-agent systems. In *Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications*, Vol. 11006. International Society for Optics and Photonics, 110060Z.
- [2] Elias Bareinboim and Judea Pearl. 2012. Transportability of causal effects: Completeness results. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 26.
- [3] Elias Bareinboim and Judea Pearl. 2013. Causal Transportability with Limited Experiments. In *Proceedings of the Twenty-Seventh National Conference on Artificial Intelligence (AAAI 2013)*, M. desjardins and M. Littman (Eds.). AAAI Press, Menlo Park, CA, 95–101.
- [4] Elias Bareinboim and Judea Pearl. 2014. Transportability from multiple environments with limited experiments: Completeness results. *Advances in neural information processing systems* 27 (2014), 280–288.
- [5] Olivier Bousquet and André Elisseeff. 2002. Stability and generalization. *The Journal of Machine Learning Research* 2 (2002), 499–526.
- [6] Glenn H Bracht and Gene V Glass. 1968. The external validity of experiments. *American educational research journal* 5, 4 (1968), 437–474.
- [7] Robert L Brennan. 1992. Generalizability theory. *Educational Measurement: Issues and Practice* 11, 4 (1992), 27–34.
- [8] Bobby J Calder, Lynn W Phillips, and Alice M Tybout. 1982. The concept of external validity. *Journal of consumer research* 9, 3 (1982), 240–244.
- [9] Yeounoh Chung, Peter J Haas, Eli Upfal, and Tim Kraska. 2018. Unknown examples & machine learning model generalization. *arXiv preprint arXiv:1808.08294* (2018).
- [10] Carlos Cinelli, Andrew Forney, and Judea Pearl. 2020. A crash course in good and bad controls. *Available at SSRN 3689437* (2020).
- [11] Andrew Forney and Elias Bareinboim. 2019. Counterfactual Randomization: Rescuing Experimental Studies from Obscured Confounding. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 2454–2461.
- [12] Andrew Forney, Judea Pearl, and Elias Bareinboim. 2017. Counterfactual Data-Fusion for Online Reinforcement Learners. In *International Conference on Machine Learning*. 1156–1164.
- [13] Kenji Kawaguchi, Leslie Pack Kaelbling, and Yoshua Bengio. 2017. Generalization in deep learning. *arXiv preprint arXiv:1710.05468* (2017).
- [14] Lawrence Leung. 2015. Validity, reliability, and generalizability in qualitative research. *Journal of family medicine and primary care* 4, 3 (2015), 324.
- [15] Lucy Ellen Lwakatare, Aiswarya Raj, Ivica Crnkovic, Jan Bosch, and Helena Holmström Olsson. 2020. Large-scale machine learning systems in real-world industrial settings: A review of challenges and solutions. *Information and Software Technology* 127 (2020), 106368.
- [16] Daniel Malinsky and David Danks. 2018. Causal discovery algorithms: A practical guide. *Philosophy Compass* 13, 1 (2018), e12470.
- [17] Karthika Mohan and Judea Pearl. 2021. Graphical models for processing missing data. *J. Amer. Statist. Assoc.* (2021), 1–16.
- [18] Andrei Paleyes, Raoul-Gabriel Urma, and Neil D Lawrence. 2020. Challenges in deploying machine learning: a survey of case studies. *arXiv preprint arXiv:2011.09926* (2020).
- [19] Judea Pearl. 1995. Causal diagrams for empirical research. *Biometrika* 82, 4 (1995), 669–688.
- [20] J. Pearl. 2009. *Causality: Models, Reasoning, and Inference* (second ed.). Cambridge University Press, New York.
- [21] Bernhard Schölkopf. 2019. Causality for machine learning. *arXiv preprint arXiv:1911.10500* (2019).
- [22] Ilya Shpitser, Tyler VanderWeele, and James M Robins. 2012. On the validity of covariate adjustment for estimating causal effects. *arXiv preprint arXiv:1203.3515* (2012).
- [23] Adarsh Subbaswamy, Peter Schulam, and Suchi Saria. 2018. Learning predictive models that transport. *arXiv preprint arXiv:1812.04597* (2018).
- [24] Richard S Sutton and Andrew G Barto. 2018. *Reinforcement learning: An introduction*. MIT press.
- [25] Victor Talpaert, Ibrahim Sobh, B Ravi Kiran, Patrick Mannion, Senthil Yogamani, Ahmad El-Sallab, and Patrick Perez. 2019. Exploring applications of deep reinforcement learning for real-world autonomous driving systems. *arXiv preprint arXiv:1901.01536* (2019).
- [26] William R Thompson. 1933. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika* 25, 3/4 (1933), 285–294.
- [27] Lisa Torrey and Jude Shavlik. 2010. Transfer learning. In *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*. IGI global, 242–264.
- [28] Guillermo Viguera and Juan A Botia. 2007. Tracking causality by visualization of multi-agent interactions using causality graphs. In *International Workshop on Programming Multi-Agent Systems*. Springer, 190–204.
- [29] Karl Weiss, Taghi M Khoshgoftaar, and DingDing Wang. 2016. A survey of transfer learning. *Journal of Big data* 3, 1 (2016), 1–40.
- [30] Kui Yu, Xindong Wu, and Hao Wang. 2010. Online causal discovery. In *9th IEEE International Conference on Cognitive Informatics (ICCI'10)*. IEEE, 667–671.