





## REFERENCES

- [1] Joshua Achiam, David Held, Aviv Tamar, and Pieter Abbeel. 2017. Constrained policy optimization. In *Proceedings of the 34th International Conference on Machine Learning (ICML)*. 22–31.
- [2] Eitan Altman. 1999. *Constrained Markov Decision Processes*. Vol. 7. CRC Press.
- [3] Kareem Amin, Nan Jiang, and Satinder Singh. 2017. Repeated inverse reinforcement learning. In *Advances in Neural Information Processing Systems (NIPS)*. 1813–1822.
- [4] Dario Amodei, Chris Olah, Jacob Steinhardt, Paul Christiano, John Schulman, and Dan Mané. 2016. Concrete problems in AI safety. *arXiv preprint arXiv:1606.06565* (2016).
- [5] Craig Boutilier, Richard Dearden, and Moisés Goldszmidt. 2000. Stochastic dynamic programming with factored representations. *Artificial intelligence* 121, 1 (2000), 49–107.
- [6] Javier Garcia and Fernando Fernández. 2015. A comprehensive survey on safe reinforcement learning. *Journal of Machine Learning Research* 16, 1 (2015), 1437–1480.
- [7] Dylan Hadfield-Menell, Smitha Milli, Stuart J Russell, Pieter Abbeel, and Anca Dragan. 2017. Inverse reward design. In *Advances in Neural Information Processing Systems (NIPS)*. 6749–6758.
- [8] Alexander Hans, Daniel Schneegaß, Anton Maximilian Schäfer, and Steffen Udfluft. 2008. Safe exploration for reinforcement learning. In *European Symposium on Artificial Neural Networks (ESANN)*. 143–148.
- [9] Michael Laskey, Sam Staszak, Wesley Yu-Shu Hsieh, Jeffrey Mahler, Florian T Pokorny, Anca D Dragan, and Ken Goldberg. 2016. SHIV: Reducing supervisor burden in DAGger using support vectors for efficient learning from demonstrations in high dimensional state spaces. In *IEEE International Conference on Robotics and Automation (ICRA)*. 462–469.
- [10] Jan Leike, Miljan Martic, Victoria Krakovna, Pedro A Ortega, Tom Everitt, Andrew Lefrancq, Laurent Orseau, and Shane Legg. 2017. AI safety gridworlds. *arXiv preprint arXiv:1711.09883* (2017).
- [11] Smitha Milli, Dylan Hadfield-Menell, Anca D. Dragan, and Stuart J. Russell. 2017. Should robots be obedient?. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI)*. 4754–4760.
- [12] Teodor M. Moldovan and Pieter Abbeel. 2012. Safe exploration in Markov decision processes. In *Proceedings of the 29th International Conference on Machine Learning (ICML)*. 1711–1718.
- [13] Arnab Nilim and Laurent El Ghaoui. 2005. Robust control of Markov decision processes with uncertain transition matrices. *Operations Research* 53, 5 (2005), 780–798.
- [14] Kevin Regan and Craig Boutilier. 2009. Regret-based reward elicitation for Markov decision processes. In *Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence (UAI)*. 444–451.
- [15] Kevin Regan and Craig Boutilier. 2010. Robust policy computation in reward-uncertain MDPs using nondominated policies. In *Association for the Advancement of Artificial Intelligence (AAAI)*. 1127–1133.
- [16] Paolo Viappiani and Craig Boutilier. 2009. Regret-based optimal recommendation sets in conversational recommender systems. In *Proceedings of the Third ACM Conference on Recommender Systems*. 101–108.
- [17] Shun Zhang, Edmund Durfee, and Satinder Singh. 2017. Approximately-optimal queries for planning in reward-uncertain Markov decision processes. In *Proceedings of the 27th International Conference on Automated Planning and Scheduling (ICAPS)*. 339–347.
- [18] Shun Zhang, Edmund Durfee, and Satinder Singh. 2018. Minimax-regret querying on side effects for safe optimality in factored Markov decision processes. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI)*. To appear.