

WILLIAM OVERMAN

655 Knight Way, Stanford, CA 94305, USA

✉ wpo@stanford.edu

💻 woverman.com

🔍 [Google Scholar](#)

in [LinkedIn](#)

💡 RESEARCH INTERESTS

AI Alignment AI Safety Reinforcement Learning Causal Inference Healthcare

🎓 EDUCATION

Stanford University	2022-2026 (<i>exp.</i>)
Graduate School of Business	
Ph.D. Operations, Information and Technology	
University of California, Irvine	2020-2022
M.Sc., Computer Science	
California Institute of Technology (Caltech)	2016-2020
B.S., Mathematics & Computer Science (Double Major)	

🏢 INDUSTRY AND RESEARCH POSITIONS

Uber	
PhD Scientist Intern, Airports	2023, Sunnyvale, CA
Institute for Basic Science	
Visiting Researcher	2021, Daejeon, South Korea
Summer Undergraduate Research Fellowship	2019, Daejeon, South Korea
Oxford University	
Summer Undergraduate Research Fellowship	2018, Oxford, United Kingdom
Mathematical Institute	

📄 PUBLICATIONS

[REFEREED CONFERENCE PAPERS]

- **The Oversight Game: Learning to Cooperatively Balance an AI Agent's Safety and Autonomy**
W. Overman, M. Bayati.
Early Version in **MLxOR Workshop** at NeurIPS 2025.
- **Conformal Arbitrage: Risk-Controlled Balancing of Competing Objectives in Language Models**
W. Overman, M. Bayati.
NeurIPS 2025: Neural Information Processing Systems, 2025.
- **Can We Validate Counterfactual Estimations in the Presence of General Network Interference?**
S. Shirani, Y. Luo, W. Overman, R. Xiong, M. Bayati.
Oral Presentation CODE@MIT 2025: Conference on Digital Experimentation @ MIT, 2025
Oral presentation at the MSOM Technology, Innovation, and Entrepreneurship SIG, 2025.
- **Aligning Model Properties via Conformal Risk Control**
W. Overman, J. Vallon, M. Bayati.
NeurIPS 2024: Neural Information Processing Systems, 2024.

- **Higher-Order Causal Message Passing for Experimentation with Complex Interference**
M. Bayati, Y. Luo, **W. Overman**, S. Shirani, R. Xiong. (alphabetical).
NeurIPS 2024: Neural Information Processing Systems, 2024.
- **Beating Price of Anarchy and Gradient Descent without Regret in Potential Games**
I. Sakos, S. Leonardos, S. Stavroulakis, **W. Overman**, I. Panageas, G. Piliouras.
ICLR 2024: International Conference on Learning Representations, 2024.
- **Global Convergence of Multi-Agent Policy Gradient in Markov Potential Games**
S. Leonardos, **W. Overman**, I. Panageas, G. Piliouras (**alphabetical, sole student**).
ICLR 2022: International Conference on Learning Representations, 2022.
- **Independent Natural Policy Gradient always converges in Markov Potential Games**
R. Fox, S. Mcaleer, **W. Overman**, I. Panageas (**alphabetical, sole student**).
AISTATS 2022. International Conference on Artificial Intelligence and Statistics, 2022

[SUBMITTED AND WORKING PAPERS]

- **Occupancy Prediction with Patient Data: Evaluating Time-Series, Patient-Level Aggregation, and Deep Set Models**
S.-H. Kim, **W. Overman**, J. Pauphilet, W. Cha. (**alphabetical, sole student**).
Major Revision at MSOM: Manufacturing & Service Operations Management.
- **On Aligning Prediction Models with Clinical Experiential Learning: A Prostate Cancer Case Study**
J. J. Vallon, **W. Overman**, W. Xu, N. Panjwani, X. Ling, S. Vij, H. P. Bagshaw, J. T. Leppert, S. Shah, G. Sonn, S. Srinivas, E. Pollom, M. K. Buyyounouski, M. Bayati.
arXiv.

[JOURNAL PAPERS]

- **Improved Regret Bound for Safe Reinforcement Learning via Tighter Cost Pessimism and Reward Optimism**
K. Yu, D. Lee, **W. Overman**, D. Lee.
Reinforcement Learning Journal, 2025.
- **Some Ordered Ramsey Numbers of Graphs on Four Vertices**
W. Overman, J. Alm, K. Coffey, C. Langhoff.
Australasian Journal of Combinatorics, 2024.

☆ AWARDS

Dean's Award, UC Irvine, for outstanding research potential. 2020
Morgan Ward Prize, Caltech, best original work in mathematics by an underclassman. 2018

✍ TEACHING

TA, Operations Stanford, Graduate School of Business, 2024 Fall
TA, Optimization and Simulation Modeling Stanford, Graduate School of Business, 2023 Fall
TA, Introduction to Algorithms University of California, Irvine 2022 Spring
TA, Introduction to Cryptography University of California, Irvine 2021 Fall
TA, Introduction to Algorithms California Institute of Technology 2020 Spring (Head TA), 2019 Spring, 2018 Spring
TA, Decidability and Tractability California Institute of Technology 2019 Fall
TA, Discrete Mathematics California Institute of Technology 2019 Winter, 2020 Winter