

1. Achiam, J., Held, D., Tamar, A. & Abbeel, P. Constrained Policy Optimization. Preprint at <https://doi.org/10.48550/arXiv.1705.10528> (2017).
2. AhmadiTeshnizi, A. *et al.* OptiMUS-0.3: Using Large Language Models to Model and Solve Optimization Problems at Scale. Preprint at <https://doi.org/10.48550/arXiv.2407.19633> (2025).
3. Amos, B. & Kolter, J. Z. OptNet: Differentiable Optimization as a Layer in Neural Networks. Preprint at <https://doi.org/10.48550/arXiv.1703.00443> (2021).
4. Bianchi, F. *et al.* How Well Can LLMs Negotiate? NegotiationArena Platform and Analysis. Preprint at <https://doi.org/10.48550/arXiv.2402.05863> (2024).
5. Boyd, S. Neal Parikh Department of Computer Science Stanford University.
6. Chen, H. *et al.* OptiChat: Bridging Optimization Models and Practitioners with Large Language Models. Preprint at <https://doi.org/10.48550/arXiv.2501.08406> (2025).
7. Chen, R., Arditi, A., Sleight, H., Evans, O. & Lindsey, J. Persona Vectors: Monitoring and Controlling Character Traits in Language Models. Preprint at <https://doi.org/10.48550/arXiv.2507.21509> (2025).
8. Chen, X., As, Y. & Krause, A. Learning Safety Constraints for Large Language Models. Preprint at <https://doi.org/10.48550/arXiv.2505.24445> (2025).
9. Chen, X., As, Y. & Krause, A. Learning Safety Constraints for Large Language Models. Preprint at <https://doi.org/10.48550/arXiv.2505.24445> (2025).
10. Cheng, R., Orosz, G., Murray, R. M. & Burdick, J. W. End-to-End Safe Reinforcement Learning through Barrier Functions for Safety-Critical Continuous Control Tasks. Preprint at <https://doi.org/10.48550/arXiv.1903.08792> (2019).
11. Constante-Flores, G. E., Chen, H. & Li, C. Enforcing Hard Linear Constraints in Deep Learning Models with Decision Rules. Preprint at <https://doi.org/10.48550/arXiv.2505.13858> (2025).
12. Dalal, G. *et al.* Safe Exploration in Continuous Action Spaces. Preprint at <https://doi.org/10.48550/arXiv.1801.08757> (2018).
13. Fang, J. *et al.* AlphaEdit: Null-Space Constrained Knowledge Editing for Language Models. Preprint at <https://doi.org/10.48550/arXiv.2410.02355> (2025).
14. Fioretto, F., Hentenryck, P. V. & Mak, T. W. K. Lagrangian Duality for Constrained Deep Learning.
15. Fioretto, F. *et al.* Lagrangian Duality for Constrained Deep Learning. Preprint at <https://doi.org/10.48550/arXiv.2001.09394> (2020).

16. Fisac, J. F. *et al.* A General Safety Framework for Learning-Based Control in Uncertain Robotic Systems. Preprint at <https://doi.org/10.48550/arXiv.1705.01292> (2018).
17. Fisac, J. F. *et al.* A General Safety Framework for Learning-Based Control in Uncertain Robotic Systems. Preprint at <https://doi.org/10.48550/arXiv.1705.01292> (2018).
18. Hentenryck, P. V. Optimization Learning. Preprint at <https://doi.org/10.48550/arXiv.2501.03443> (2025).
19. Herbert, S. *et al.* Scalable Learning of Safety Guarantees for Autonomous Systems using Hamilton-Jacobi Reachability. Preprint at <https://doi.org/10.48550/arXiv.2101.05916> (2021).
20. Jiang, H. & Haghtalab, N. On Surjectivity of Neural Networks: Can you elicit any behavior from your model? Preprint at <https://doi.org/10.48550/arXiv.2508.19445> (2025).
21. Jimenez, C. E. *et al.* SWE-bench: Can Language Models Resolve Real-World GitHub Issues? Preprint at <https://doi.org/10.48550/arXiv.2310.06770> (2024).
22. Leung, K., Aréchiga, N. & Pavone, M. Backpropagation through Signal Temporal Logic Specifications: Infusing Logical Structure into Gradient-Based Methods. Preprint at <https://doi.org/10.48550/arXiv.2008.00097> (2021).
23. Min, Y. & Azizan, N. HardNet: Hard-Constrained Neural Networks with Universal Approximation Guarantees. Preprint at <https://doi.org/10.48550/arXiv.2410.10807> (2025).
24. Min, Y. & Azizan, N. HardNet: Hard-Constrained Neural Networks with Universal Approximation Guarantees. Preprint at <https://doi.org/10.48550/arXiv.2410.10807> (2025).
25. Patwardhan, T. *et al.* GDPval: Evaluating AI Model Performance on Real-World Economically Valuable Tasks. Preprint at <https://doi.org/10.48550/arXiv.2510.04374> (2025).
26. Ramirez, J., Hashemizadeh, M. & Lacoste-Julien, S. Position: Adopt Constraints Over Penalties in Deep Learning. Preprint at <https://doi.org/10.48550/arXiv.2505.20628> (2025).
27. Santos, H. V. F., Costa, V., Montandon, J. E. & Valente, M. T. Decoding the Configuration of AI Coding Agents: Insights from Claude Code Projects. Preprint at <https://doi.org/10.48550/arXiv.2511.09268> (2025).
28. Simchi-Levi, D., Dai, T., Menache, I. & Wu, M. X. Democratizing Optimization with Generative AI.
29. Simchi-Levi, D., Mellou, K., Menache, I. & Pathuri, J. Large Language Models for Supply Chain Decisions. Preprint at <https://doi.org/10.48550/arXiv.2507.21502> (2025).
30. Sinha, R. *et al.* Real-Time Anomaly Detection and Reactive Planning with Large Language Models. Preprint at <https://doi.org/10.48550/arXiv.2407.08735> (2024).

31. Varambally, S. *et al.* Hilbert: Recursively Building Formal Proofs with Informal Reasoning. Preprint at <https://doi.org/10.48550/arXiv.2509.22819> (2025).
32. Xiao, W. *et al.* Differentiable Control Barrier Functions for Vision-based End-to-End Autonomous Driving. Preprint at <https://doi.org/10.48550/arXiv.2203.02401> (2022).
33. Xiao, W. *et al.* Differentiable Control Barrier Functions for Vision-based End-to-End Autonomous Driving. Preprint at <https://doi.org/10.48550/arXiv.2203.02401> (2022).
34. Xiao, W. *et al.* BarrierNet: Differentiable Control Barrier Functions for Learning of Safe Robot Control. *IEEE Transactions on Robotics* **39**, 2289–2307 (2023).
35. Xiao, W. *et al.* BarrierNet: Differentiable Control Barrier Functions for Learning of Safe Robot Control. *IEEE Transactions on Robotics* **39**, 2289–2307 (2023).
36. Yang, T.-Y., Rosca, J., Narasimhan, K. & Ramadge, P. J. Projection-Based Constrained Policy Optimization. Preprint at <https://doi.org/10.48550/arXiv.2010.03152> (2020).
37. Zhang, Y., Vuong, Q. & Ross, K. W. First Order Constrained Optimization in Policy Space. Preprint at <https://doi.org/10.48550/arXiv.2002.06506> (2020).
38. Zhang, Z. *et al.* Distributionally Robust Constrained Reinforcement Learning under Strong Duality. Preprint at <https://doi.org/10.48550/arXiv.2406.15788> (2024).
39. Advancing LLM Safe Alignment with Safety Representation Ranking. in (2025).
40. Lin, A., Peng, S. & Bansal, S. One Filter to Deploy Them All: Robust Safety for Quadrupedal Navigation in Unknown Environments. Preprint at <https://doi.org/10.48550/arXiv.2412.09989> (2024).
41. Guiding LLMs The Right Way: Fast, Non-Invasive Constrained Generation. <https://arxiv.org/html/2403.06988v1>.
42. Chen, Z. *et al.* OptiMind: Teaching LLMs to Think Like Optimization Experts. (2025).
43. Claude Code Best Practices. <https://www.anthropic.com/engineering/clause-code-best-practices>.