Towards Safe Robot Foundation Models

Maximilian Tölle*,[1,2] Theo Gruner*,[1,3] Daniel Palenicek*,[1,3] Jonas Günster*,[1]
Puze Liu[1,2] Joe Watson[4] Davide Tateo[1] Jan Peters[1,2,3,5,6]

***Abstract*— Robot foundation models hold the potential for deployment across diverse environments, from industrial applications to household tasks. While current research focuses primarily on the policies' generalization capabilities across a variety of tasks, it fails to address safety, a critical requirement for deployment on real-world systems. In this paper, we introduce a safety layer designed to constrain the action space of any generalist policy appropriately. Our approach uses ATACOM, a safe reinforcement learning algorithm that creates a safe action space and, therefore, ensures safe state transitions. By extending ATACOM to generalist policies, our method facilitates their deployment in safety-critical scenarios without requiring any specific safety fine-tuning. We demonstrate the effectiveness of this safety layer in an air hockey environment, where it prevents a puck-hitting agent from colliding with its surroundings, a failure observed in generalist policies. https://sites.google.com/robot-learning.de/towards-safe-rfm**

I. Introduction

Deploying autonomous agents in real-world environments requires motion generation that is both feasible and adaptable to various scenarios. Robot foundation models (RFMs) advance this goal by being trained across a variety of embodiments, tasks, and environments. However, a critical component—safety—remains unaddressed despite its importance in many real-world applications. Current RFMs [1], [2] are typically trained with behavior cloning (BC) to imitate expert trajectories. Given that expert data predominantly consists of safe demonstrations, RFMs may implicitly reflect a notion of safety as a result of this data bias. However, while this may encourage conservative behavior in safety-critical tasks, it does not provide any formal safety guarantees. Additionally, BC policies may catastrophically damage the robot during deployment when encountering unseen observations due to the distribution shift [3], [4], [5]. Therefore, we argue that integrating domain expertise is essential for ensuring reliable safety.

Inductive biases combat several shortcomings of purely data-driven approaches in robot learning [6], [7], [8], [9], [10]. Incorporating analytic models into these optimization techniques enables sound inference [10], data-efficient learning [9], and safety [6] by exploiting the problem's inherent dynamic structure. Safety has thereby been a major concern within the control community which has developed several safety-ensuring methods using control barrier functions [11], [12], [13], [14], [15], reachability analysis [16], [17], [18], [19], [20] and shielding [21], [22], [23], [24]. Commonly, all these approaches exploit domain knowledge to construct a guaranteed safety filter. For more details, refer to the following reviews [25], [26], [27].

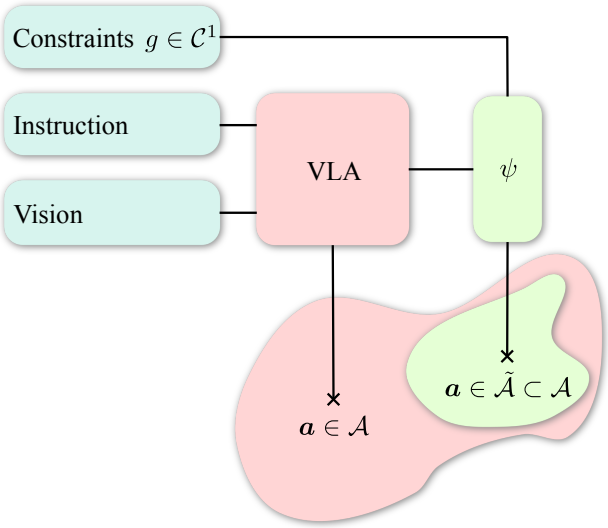

Fig. 1. Composition of a safe VLA policy. The proposed safety module ψ depends on a pre-specified set of safety constraints g and can be added on top of the output of the VLA policy. While the initial output of the VLA policy does not guarantee safe actions, the added safety module ensures safe state transitions.

In this paper, we introduce a safety module that enables a pre-trained RFM to operate safely within an environment by adhering to domain-specific safety constraints. This is accomplished by utilizing system dynamics to control actions within the robot's null space. Following [6], [28], we create a safe action space from the constraints and system dynamics, which ensures that an initially unsafe action from a RFM is mapped to a safe action. By doing this, the module ensures safe transitions during deployment.

II. A Safety Module For Generalist Policies

Today's RFMs [29], [30], [1], [2] are trained on large-scale datasets [30] to predict actionable outputs from multi-modal observations $\boldsymbol{x}$, such as images, language instructions and proprioceptive data. We define the policy that maps language instruction and observations to robotic actions as

This work was supported by "Third Wave of AI", funded by the Excellence Program of the Hessian Ministry of Higher Education, Science, Research and Art. We also acknowledge the grant "Einrichtung eines Labors des Deutschen Forschungszentrum für Künstliche Intelligenz (DFKI) an der Technischen Universität Darmstadt" of the Hessian Ministry of Science and Research, Arts and Culture.

*Equal contribution [1]Technical University of Darmstadt [2]German Research Center for Artificial Intelligence (DFKI) [3]hessian.AI [4]University of Oxford [5]Robotics Institute Germany (RIG) [6]Centre for Cognitive Science
Correspondence:
{maximilian,theo,palenicek}@robot-learning.de

$\boldsymbol{a} \sim \pi_{\text{VLA}}(\cdot \mid \boldsymbol{x})$. Our goal is to make an already trained RFM safe at test time. Therefore, we adopt ATACOM [28] to ensure guaranteed safe actions. We pose the following requirements for the system:

Requirement 1: Access to the system's state $\boldsymbol{s}$ and a control affine system $\dot{\boldsymbol{s}} = f(\boldsymbol{s}) + G(\boldsymbol{s})\boldsymbol{a}$.

Requirement 2: We define the safety conditions as continuously differentiable constraints $\mathbf{0} \geq g(\boldsymbol{x}) \in \mathcal{C}^1$.

Most robotic manipulators in [30] fulfill rigid-body assumptions and thus already comply with Requirement 1. Furthermore, we assume that practitioners have prior knowledge of the robot's safety requirements and can effectively define the system's constraints (Requirement 2). Thus, while the above-stated requirements may seem restrictive at first, we deem that these assumptions hold for most currently considered robotic platforms VLAs are trained on.

Acting on the tangent space of the constraint manifold. Building on the aforementioned requirements, ATACOM [6], [28] constructs a constraint manifold of safe configurations. Actions are then mapped into the tangent space of this manifold, ensuring safe transitions. As such, ATACOM can be seen as a mapping from actions $\boldsymbol{a}$, the state $\boldsymbol{s}$, and the safety constraints g to safe actions

$$\boldsymbol{a}_{\text{safe}} = \psi_{G,f}(\boldsymbol{a}, \boldsymbol{s}, g), \quad \boldsymbol{a} \sim \pi_{\text{VLA}}(\cdot \mid \boldsymbol{x}).$$

In this way, we ensure that actions that are drawn from a VLA policy are mapped to be safe actions that guarantee compliance with the safety constraints g.

III. Experiments

We empirically evaluate the proposed approach on a robot air hockey task. The objective is to hit a puck into the goal while adhering to multiple safety constraints, such as keeping the end-effector on the table surface, preventing the arm from colliding with the table, and ensuring joint position limits. We refer to [28] for a detailed description of the experimental setup. The policy's observation consists of language instructions, a goal image of the scene, and proprioceptive data in the form of joint positions, joint velocities, puck position, and puck velocity. While not needed for safety but for improved performance in the air hockey task, we fine-tune a pre-trained OCTO [1] policy using behavior cloning in both a simulated MUJOCO [31] environment and a real-world setting. Importantly, we obtain the fine-tuning data by an expert policy that does not leverage ATACOM. The policy outputs desired end-effector velocities in the x-y plane of the table surface, which are converted to joint velocities using inverse kinematics. The ATACOM layer then maps these joint velocities to safe ones before passing them to a joint-space controller. We compare our safety-aware approach to an unsafe baseline, where the joint-space controller directly executes the unfiltered joint velocities.

We evaluate the safety module for various fine-tuning checkpoints of OCTO on the physical system. Several deployment videos of OCTO playing air hockey can be found on our project page. Fig. 2 shows that the OCTO agent with the added safety module does not violate the safety constraints during inference. On the contrary, OCTO without the added safety layer heavily violates the constraints even though the fine-tuning data contains safe expert demonstrations. Looking at the success rates, we observe a steady performance improvement in the number of training iterations when using the safety module. Importantly, while the fine-tuning data is not obtained with ATACOM, we still obtain high success rates, which underlines that ATACOM does not generate overly conservative control actions. Interestingly, we see that the constraint violations of the OCTO baseline increase with added training time, which negatively impacts the policy performance.

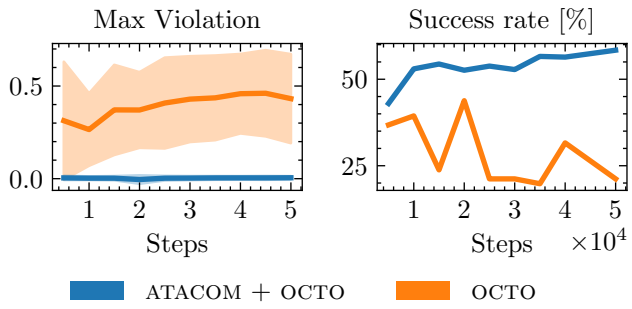

Fig. 2. Safety violations of the OCTO policy w/o the safety module on the air hockey hitting task for different checkpoints during the training phase. We report the maximum constraint violation a trajectory as well as the success rate of the robot hitting the puck into the goal. When the ATACOM safety module is added, the policy remains compliant with safety constraints throughout fine-tuning. It progressively improves its success rate, whereas the unmodified OCTO policy continues to breach safety limits while achieving a lower success rate.

IV. Conclusion

We propose a safety module that can be added as the final layer of a Robot foundation model (RFM) by leveraging domain-specific knowledge. Although leveraging domain knowledge may seem counterintuitive for RFMs, we hypothesize that reasoning with system dynamics is essential for ensuring safety. Additionally, by designing this module as an independent safety layer, it does not incorporate any additional computational burden, such as fine-tuning, to ensure safety. We demonstrate the effectiveness of the safety layer by evaluating a VLA policy with BC on an air hockey hitting task for which it is critical not to crash with the tabletop. While we emphasize that ensuring safety requires domain expertise, it can also be a demanding task to formulate all scene-specific safety constraints. One intuitive research direction is to automate the process by leveraging the inherent knowledge of vision-language models (VLMs). However, so far, VLMs have only been used to integrate semantic safety constraints such as "keep the cup upright" into an already existing set of constraints [32], [33]. Beyond the formulation of safety constraints, it remains an open research question of how a more generalizable concept of safety can be formulated and applied across different embodiments, environments, and tasks.

REFERENCES

[1] Octo Model Team, D. Ghosh, H. Walke, K. Pertsch, K. Black, O. Mees, *et al.*, "Octo: An open-source generalist robot policy," in *Proceedings of Robotics: Science and Systems*, 2024.

[2] M. J. Kim, K. Pertsch, S. Karamcheti, T. Xiao, A. Balakrishna, S. Nair, *et al.*, "OpenVLA: An Open-Source Vision-Language-Action Model," in *8th Conference on Robot Learning (CoRL)*, 2024.

[3] J. A. Bagnell, "An invitation to imitation," *Technical Report*, 2015.

[4] T. Osa, J. Pajarinen, G. Neumann, J. A. Bagnell, P. Abbeel, J. Peters, *et al.*, "An algorithmic perspective on imitation learning," *Foundations and Trends® in Robotics*, vol. 7, no. 1-2, pp. 1–179, 2018.

[5] S. Ross, G. Gordon, and D. Bagnell, "A reduction of imitation learning and structured prediction to no-regret online learning," in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, ser. Proceedings of Machine Learning Research, vol. 15, 2011, pp. 627–635.

[6] P. Liu, D. Tateo, H. B. Ammar, and J. Peters, "Robot reinforcement learning on the constraint manifold," in *5th Conference on Robot Learning (CoRL*. PMLR, 2021.

[7] T.-W. Ke, N. Gkanatsios, and K. Fragkiadaki, "3d diffuser actor: Policy diffusion with 3d scene representations," in *8th Annual Conference on Robot Learning*, 2024.

[8] N. Funk, J. Urain, J. Carvalho, V. Prasad, G. Chalvatzaki, and J. Peters, "Actionflow: Equivariant, accurate, and efficient policies with spatially symmetric flow matching," *arXiv preprint arXiv:2409.04576*, 2024.

[9] M. Lutter, C. Ritter, and J. Peters, "Deep lagrangian networks: Using physics as model prior for deep learning," in *International Conference on Learning Representations (ICLR)*, 2019.

[10] F. Muratore, T. Gruner, F. Wiese, B. Belousov, M. Gienger, and J. Peters, "Neural posterior domain randomization," in *5th Conference on Robot Learning (CoRL)*, 2021.

[11] A. D. Ames, S. Coogan, M. Egerstedt, G. Notomista, K. Sreenath, and P. Tabuada, "Control barrier functions: Theory and applications," in *2019 18th European control conference (ECC)*. IEEE, 2019, pp. 3420–3431.

[12] A. Taylor, A. Singletary, Y. Yue, and A. Ames, "Learning for safety-critical control with control barrier functions," in *Learning for Dynamics and Control*. PMLR, 2020, pp. 708–717.

[13] W. Xiao and C. Belta, "High-Order Control Barrier Functions," *IEEE Transactions on Automatic Control*, vol. 67, no. 7, pp. 3655–3662, July 2022, conference Name: IEEE Transactions on Automatic Control.

[14] D. C. Tan, F. Acero, R. McCarthy, D. Kanoulas, and Z. A. Li, "Your value function is a control barrier function: Verification of learned policies using control theory," *arXiv preprint arXiv:2306.04026*, 2023.

[15] Y. Yang, Y. Jiang, Y. Liu, J. Chen, and S. E. Li, "Model-free safe reinforcement learning through neural barrier certificate," *IEEE Robotics and Automation Letters*, vol. 8, no. 3, pp. 1295–1302, 2023.

[16] A. K. Akametalu, J. F. Fisac, J. H. Gillula, S. Kaynama, M. N. Zeilinger, and C. J. Tomlin, "Reachability-based safe learning with gaussian processes," in *53rd IEEE conference on decision and control*. IEEE, 2014, pp. 1424–1431.

[17] J. F. Fisac, A. K. Akametalu, M. N. Zeilinger, S. Kaynama, J. Gillula, and C. J. Tomlin, "A general safety framework for learning-based control in uncertain robotic systems," *IEEE Transactions on Automatic Control*, vol. 64, no. 7, pp. 2737–2752, 2018.

[18] Y. S. Shao, C. Chen, S. Kousik, and R. Vasudevan, "Reachability-based trajectory safeguard (rts): A safe and fast reinforcement learning safety layer for continuous control," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 3663–3670, 2021.

[19] M. Selim, A. Alanwar, S. Kousik, G. Gao, M. Pavone, and K. H. Johansson, "Safe Reinforcement Learning Using Black-Box Reachability Analysis ," *IEEE Robotics and Automation Letters*, 2022.

[20] K. P. Wabersich, A. J. Taylor, J. J. Choi, K. Sreenath, C. J. Tomlin, A. D. Ames, and M. N. Zeilinger, "Data-driven safety filters: Hamilton-jacobi reachability, control barrier functions, and predictive methods for uncertain systems," *IEEE Control Systems Magazine*, vol. 43, no. 5, pp. 137–177, 2023.

[21] M. Alshiekh, R. Bloem, R. Ehlers, B. Könighofer, S. Niekum, and U. Topcu, "Safe reinforcement learning via shielding," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 32, no. 1, 2018.

[22] G. Dalal, K. Dvijotham, M. Vecerik, T. Hester, C. Paduraru, and Y. Tassa, "Safe exploration in continuous action spaces," *arXiv preprint arXiv:1801.08757*, 2018.

[23] L. Hewing, K. P. Wabersich, M. Menner, and M. N. Zeilinger, "Learning-based model predictive control: Toward safe learning in control," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 3, no. 1, pp. 269–296, 2020.

[24] S. Carr, N. Jansen, S. Junges, and U. Topcu, "Safe reinforcement learning via shielding under partial observability," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 12, 2023, pp. 14 748–14 756.

[25] L. Brunke, M. Greeff, A. W. Hall, Z. Yuan, S. Zhou, J. Panerati, and A. P. Schoellig, "Safe learning in robotics: From learning-based control to safe reinforcement learning," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 5, no. Volume 5, 2022, pp. 411–444, 2022.

[26] K. P. Wabersich, A. J. Taylor, J. J. Choi, K. Sreenath, C. J. Tomlin, A. D. Ames, and M. N. Zeilinger, "Data-driven safety filters: Hamilton-jacobi reachability, control barrier functions, and predictive methods for uncertain systems," *IEEE Control Systems Magazine*, vol. 43, no. 5, pp. 137–177, 2023.

[27] K.-C. Hsu, H. Hu, and J. F. Fisac, "The safety filter: A unified view of safety-critical control in autonomous systems," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 7, no. Volume 7, 2024, pp. 47–72, 2024.

[28] P. Liu, H. Bou-Ammar, J. Peters, and D. Tateo, "Safe reinforcement learning on the constraint manifold: Theory and applications," *arXiv preprint arXiv:2404.09080*, 2024.

[29] A. Brohan, N. Brown, J. Carbajal, Y. Chebotar, X. Chen, K. Choromanski, *et al.*, "RT-2: Vision-Language-Action Models Transfer Web Knowledge to Robotic Control," in *7th Conference on Robot Learning (CoRL)*, 2023.

[30] O. X.-E. Collaboration, A. O'Neill, A. Rehman, A. Gupta, A. Maddukuri, A. Gupta, *et al.*, "Open X-Embodiment: Robotic learning datasets and RT-X models," https://arxiv.org/abs/2310.08864, 2023.

[31] E. Todorov, T. Erez, and Y. Tassa, "Mujoco: A physics engine for model-based control," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2012, pp. 5026–5033.

[32] L. Santos, Z. Li, L. Peters, S. Bansal, and A. Bajcsy, "Updating robot safety representations online from natural language feedback," 2024. [Online]. Available: https://arxiv.org/abs/2409.14580

[33] L. Brunke, Y. Zhang, R. Römer, J. Naimer, N. Staykov, S. Zhou, and A. P. Schoellig, "Semantically safe robot manipulation: From semantic scene understanding to motion safeguards," 2024. [Online]. Available: https://arxiv.org/abs/2410.15185