Position: Train Robustly and Certify at Runtime Under Partial Observability

Can Zhou
Department of EEE
Imperial College London
London, UK
c.zhou24@imperial.ac.uk

S. Karthik Mukkavilli
Mercuria Energy Group
Geneva, Switzerland
Korea Advanced Institute of Science and Technology
drkarthik@kaist.ac.kr

Yulong Gao
Department of EEE
Imperial College London
London, UK
yulong.gao@imperial.ac.uk

***Abstract*—AI systems are increasingly deployed in safety-critical physical environments, where actions taken under uncertainty can have irreversible real-world consequences. This position paper argues that Safe Physical AI should be built around a unified uncertainty-aware pipeline that couples robust modelling, robust policy synthesis, and runtime certification. We take the position that robust partially observable Markov decision processes provide a natural modelling abstraction for hidden state, noisy observations, uncertain dynamics, and deployment-time distribution shift. Under this formulation, policies should combine worst-case performance with probabilistic safety constraints across plausible models, while proposed actions should be certified online before execution. We further argue that runtime verifier feedback can be used to refine robust policies and uncertainty sets, closing the loop between deployment and training. Finally, we discuss how the same train-certify-refine architecture can support the safe deployment of generative AI in physical systems. This paper further identifies key open problems towards building Safe Physical AI systems that remain reliable under incomplete information and real-world uncertainty.**

***Index Terms*—Cyber-physical systems, robust partially observable Markov decision processes, formal verification.**

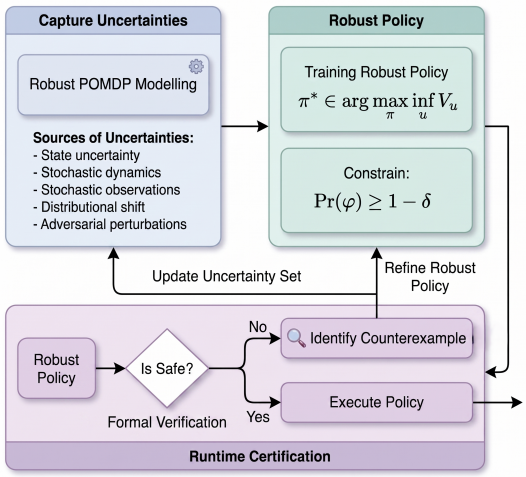

Fig. 1. Our proposed safe physical AI pipeline combining robust POMDP modelling, robust policy training, and runtime certification with counterexample-guided refinement.

I. Introduction

AI systems are increasingly deployed in domains where decisions directly affect the physical world, including autonomous vehicles, mobile robots, and industrial cyber-physical systems [1]–[3]. Operating in dynamic and safety-critical environments, these systems execute actions that directly alter real-world states, where failures may lead to irreversible consequences such as collisions, unsafe human–robot interactions, or loss of operational control [4]–[8]. Consequently, safety cannot be treated as a secondary optimisation objective. Rather, it must instead serve as a foundational principle governing uncertainty representation, policy learning, and action evaluation before deployment.

A key challenge is that real-world cyber-physical systems (CPS) operate under multiple sources of uncertainty, including partial observability by noisy sensors, uncertain dynamics with distribution shift, and unpredictable human behaviour. In many settings, the true physical state cannot be directly observed, requiring the agent to make decisions from incomplete and noisy observations. For example, a robot may not know the exact properties of an object before interaction, while an autonomous vehicle may encounter occluded pedestrians or uncertain driver behaviour. Unsafe behaviour can therefore arise not only from incorrect actions, but also from incorrect or overconfident estimates of the environment. For these reasons, modelling Physical AI systems as fully observable decision processes is often unrealistic. We argue that robust partially observable Markov decision processes (robust POMDPs) provide a more suitable framework for safety-critical cyber-physical systems because they explicitly model hidden state and uncertainty through stochastic dynamics, enabling decision-making under incomplete information.

However, modelling uncertainty alone is insufficient for safe deployment. Once uncertainty is explicitly represented, it should guide both robust policy synthesis during training and runtime safety assurance during deployment. During training, policies should be synthesised to remain robust under partial observability, and distribution shift in stochastic dynamics. A robust policy should therefore optimise behaviour not only for nominal conditions, but across plausible hidden states and uncertain environment dynamics. Robust training is important because runtime safety mechanisms alone may repeatedly reject unsafe actions without producing a policy that behaves

safely and effectively under uncertainty.

Beyond robust training, we argue that runtime certification is also required to verify whether the action to be executed remains safe under the real-time uncertainty estimate and operating conditions. This is vital since real-world environments inevitably contain situations that were not fully represented during training. If safety cannot be certified, the system should switch to a conservative fallback controller, gather additional information, request human intervention, or stop execution. Runtime certification therefore acts as a final safeguard against out-of-distribution conditions and unsafe execution.

Position: We argue against relying on robust policy synthesis or runtime certification in isolation. Safe Physical AI should instead use a unified uncertainty-aware pipeline: model the system as a robust POMDP, synthesise policies that combine worst-case performance with probabilistic safety constraints under partial observability and uncertain dynamics, certify proposed actions during execution, and use verifier feedback for subsequent refinement.

II. Modelling as a Robust POMDP

In this section, we formalise robust POMDPs as a mathematical framework for modelling safe physical AI systems. We show that this formulation can capture multiple sources of uncertainty, including uncertainty in system dynamics, observations, and deployment conditions. We then use this framework to identify open problems in uncertainty-aware safety reasoning for CPS.

A. Robust POMDPs

Definition 1 (Robust POMDP)**.** *A robust partially observable Markov decision process is a tuple* $M = \langle S, A, O, T, Z, U, R, \rho_0, \gamma \rangle$ *, where:*

- *S, A, and O are finite sets of states, actions, and observations;*
- *$T = \{T_u : S \times A \to \Delta(S)\}$ is the set of transition kernels;*
- *$Z = \{Z_u : S \times A \times S \to \Delta(O)\}$ is the set of observation kernels;*
- *$U \subseteq T \times Z$ is an uncertainty set over transition and observation kernels;*
- *$R : S \times A \to \mathbb{R}$ is the reward function;*
- *$\rho_0 \in \Delta(S)$ is the initial state distribution;*
- *$\gamma \in [0, 1)$ is the discount factor.*

Each element $u = (T_u, Z_u) \in U$ specifies one possible deployment model, where $T_u(s' \mid s, a)$ is the probability of transitioning to state s' after taking action a in state s, and $Z_u(o \mid s, a, s')$ is the probability of observing o after the transition (s, a, s'). These models are related through standard special cases. A robust POMDP with a singleton uncertainty set reduces to a standard POMDP. Under full observability, a robust POMDP reduces to a robust MDP [9]; with a singleton uncertainty set, a robust MDP reduces to a standard MDP [10].

Let Π denote a chosen class of admissible observation-based policies, such as belief-state policies [11] or finite-state-controller policies [12]. For a fixed model $u = (T_u, Z_u) \in U$, let $V_u(\pi)$ be the expected discounted return of policy $\pi \in \Pi$. The robust value of π is $V_U(\pi) = \inf_{u \in U} V_u(\pi)$. ptimal robust value is $V_U^\star = \sup_{\pi \in \Pi} \inf_{u \in U} V_u(\pi)$, and, when this supremum is attained, an optimal robust policy satisfies $\pi^\star \in \arg\max_{\pi \in \Pi} \inf_{u \in U} V_u(\pi)$.

B. Capturing Uncertainty with Robust POMDPs

Robust POMDPs provide a suitable framework for modelling safe physical AI, as the uncertainty set U can capture distributional shifts in both system dynamics and observations. Specifically, uncertainty in the transition kernel T_u represents changes in dynamics, such as variations in environmental conditions, while uncertainty in the observation kernel Z_u captures changes in sensing, including noise and perception errors. Rather than assuming a single nominal deployment model, a robust POMDP therefore considers a family of plausible models indexed by $u \in U$. This is particularly appropriate for safe physical AI, where training and deployment conditions may differ and safety must be assessed under worst-case and adverse shifts in dynamics, observations.

While robust POMDPs provide a generalised framework for capturing sources of uncertainty that standard POMDPs and robust MDPs cannot fully represent, a central challenge is constructing suitable robust POMDP models for safe physical AI reasoning. In particular, the definitions of uncertainty sets and policy classes are critical for both reasoning and verification. As discussed in [13], [14], different uncertainty-set structures, such as interval, convex, and non-convex sets, can induce optimisation problems that differ substantially in complexity, scalability, and solvability. Likewise, the choice of policy class is consequential: belief-based policy synthesis can differ fundamentally from finite-memory controller synthesis in both theoretical properties and practical computation [11], [12], [15]. Existing approaches provide limited systematic guidance on how to select uncertainty sets and policy classes for safe physical AI reasoning. Hence, this position paper argues for the following open problems.

Open Problems:

- How should uncertainty sets be defined for robust POMDPs in safe physical AI reasoning?
- Which policy classes are suitable for robust POMDP-based reasoning in safe physical AI?

III. Train Robustly and Certify at Runtime

This section details the realisation of our proposed unified pipeline grounded in the reasoning over the robust POMDP. Particularly, we move from an offline phase that optimises for general resilience to an online phase that enforces certified safety, finally closing the loop through a feedback mechanism that refines the agent's behaviour based on real-world safety interventions. The central claim is that safe physical AI

requires all three components: robust offline synthesis, formal runtime certification, and continual feedback from deployment.

A. Offline Synthesis: Learning to be Robust

In safe physical AI, offline synthesis must account for the fact that the training model is only an approximation of the deployment environment [16]. Under distribution shift, the realised transition and observation kernels may differ from the nominal model, so a policy that appears safe during training may behave unreliably after deployment [17]. The objective is therefore to synthesise policies whose performance and safety are robust to plausible model errors.

Let $u_0 = (T_{u_0}, Z_{u_0})$ denote the nominal model, and let $u = (T_u, Z_u) \in U$ denote a plausible deployment model within an uncertainty set U. The sensitivity of a policy π to model shift can be measured by $\Delta_U(\pi) = V_{u_0}(\pi) - \inf_{u \in U} V_u(\pi)$. A large $\Delta_U(\pi)$ indicates that the policy performs well under the nominal model but degrades substantially under adverse dynamics or observations. Robust offline synthesis should therefore prefer policies with high worst-case value and low sensitivity to model variation. Practically, $\Delta_U(\pi)$ can be estimated over sampled or extremal models by simulation or model checking.

Under partial observability, model sensitivity arises not only from changes in physical dynamics, but also from changes in observations and belief updates. Robustness must therefore be considered at the level of observation-based decision-making, rather than only at the level of fully observable state-action choices. Moreover, safety can be incorporated by imposing constraints during offline synthesis. For example, one may require $\Pr_u^\pi(\varphi) \geq 1 - \delta, \quad \forall u \in U$, where φ is a safety specification and δ is an acceptable risk threshold. Such a constraint provides a policy-level certificate under the assumed uncertainty set.

Existing work has studied robust POMDP synthesis, where policies are optimised against uncertainty in transition or observation models [11], [18]–[20]. This is useful for reasoning about distribution shift, but synthesis and evaluation are already computationally difficult under partial observability, and rich uncertainty sets [14]. Separately, constrained POMDPs encode safety or resource requirements through constraint costs and admissible thresholds [21]–[23]. Related work has also explored intersections between robustness, formal guarantees, and safety, including robust synthesis under formal specifications [12] and shielding for POMDPs [24], [25].

Safe physical AI requires these directions to be integrated: policies must remain robust to model uncertainty while satisfying safety constraints. This motivates robust constrained POMDPs, in which performance and constraint satisfaction must hold across a set of plausible deployment models. However, unrestricted optimal synthesis is undecidable in general for infinite-horizon POMDPs [26], while robustness and constraints introduce additional worst-case and feasibility requirements. A credible path towards tractability is therefore to pursue sound approximations, such as certified value under-approximations [25], [27], finite memory controllers [12], conformal safety shielding [28], or sampling-based search [29]. The central challenge is to determine which combinations of these techniques can scale while preserving robust constraint guarantees and meaningful performance bounds. This leads to the following open problem.

Open Problem:

- How can robust constrained POMDP policies be synthesised efficiently under both model uncertainty and partial observability with soundness guarantee?

B. Runtime Certification: Formal Verification

Robust offline synthesis can reduce unsafe behaviour by optimising policies against a family of plausible deployment models, but it cannot fully guarantee safety at deployment time: the true environment may fall outside the assumed uncertainty set, rare safety-critical events may be absent from the training data, and partial observability may cause the agent to act on an incorrect or overconfident belief. Consequently, even a robust policy may propose actions that are unsafe under the current deployment conditions. Runtime verification therefore provides a final safety layer by certifying a proposed action before it is executed on the physical system.

Certification. A key design question is what form of certificate is appropriate for safe physical AI under partial observability and deployment-time uncertainty. In some CPS, safety requires hard guarantees (almost-sure certificates) [30], where unsafe states must be avoided for all states and models consistent with the current uncertainty estimate (belief information b_t), e.g., $\Pr_u^\pi(\varphi \mid b_t, a_t) = 1, \ \forall u \in U_t$. In other settings, hard guarantees may be too conservative or infeasible due to noisy sensing, stochastic dynamics, and incomplete information, so certification may instead rely on probabilistic requirements [31]–[33], such as robust chance constraints $\inf_{u \in U_t} \Pr_u^\pi(\varphi \mid b_t, a_t) \geq 1 - \delta$, or marginal-probability constraints over beliefs. The choice of certificate determines both the strength of the safety guarantee and the computational cost of verification.

It also raises the question of what information the verifier should use: certification may rely on the uncertainty set U used during robust training, giving a consistent training-deployment pipeline, or it may construct a deployment-time uncertainty set U_t directly from online observations, state estimation [34], conformal prediction [28], [35], or belief-support tracking [29], [36]. Designing runtime certification therefore requires jointly choosing the safety requirement and the source of uncertainty information used to verify it.

Approaches. Several techniques can support runtime certification and shielding. Control barrier functions [37]–[39] can enforce forward invariance of safe sets in dynamic systems [40]–[42], reachability analysis [43]–[47] can assess whether unsafe states are reachable under uncertain dynamics or disturbances, and formal-methods-based shields [24], [48], [49] can override learned-policy actions that violate temporal constraints.

The main challenge is scalability [50]. Formal verification can become computationally expensive for realistic CPS with continuous state spaces, nonlinear dynamics, high-dimensional observations, and rich uncertainty sets. This difficulty is amplified in belief space, where verification must reason over hidden-state distributions and possibly multiple transition and observation models. As a result, strong certificates may be too conservative or too slow for online deployment, while scalable approximate shields may provide weaker guarantees. Runtime certification therefore requires the certificate and verification technique to be designed together, balancing guarantee strength with online tractability.

We therefore argue the following open problems:

Open Problems:

- What form of runtime certification is appropriate for safe physical AI under partial observability and deployment-time model uncertainty?
- How can runtime shields and verifiers scale to realistic CPS while preserving meaningful safety guarantees?

C. Closing the Loop: Feedback and Refinement

Runtime certification should not only prevent unsafe execution, but also provide feedback for improving the policy, uncertainty model, and verifier. When a verifier rejects a proposed action, the rejection identifies a deployment scenario in which the learned policy cannot be certified as safe. Such counterexamples [51]–[53] highlight regions of the state or belief space where the policy is unsafe, overconfident, or poorly aligned with the assumptions used during offline synthesis.

These counterexamples can be fed back into the robust POMDP pipeline. They may be used to improve policy synthesis in safety-critical regions, update the uncertainty set U when deployment observations reveal unmodelled dynamics or sensing errors, or refine the certification rule when the verifier is too conservative or too weak. In this way, runtime verification closes the loop between deployment and training: robust synthesis produces an initial policy, runtime certification detects unsafe or uncertain actions, and counterexamples from deployment guide subsequent refinement.

However, feedback-based refinement must preserve safety guarantees. Counterexamples may be sparse, biased towards verifier interventions, or expensive to validate on physical systems. Moreover, updating the policy, uncertainty set, or verifier may invalidate previous certificates. Each refinement should therefore remain a candidate offline update: the deployed certificate stays tied to the previous policy, U, and verifier until the revised version is re-certified. Counterexamples may enlarge U; shrinking it requires independent validation. The previous controller and fallback remain active during this process. This motivates the following open problems:

Open Problems:

- How can verifier-generated counterexamples be used to refine robust POMDP policies, uncertainty sets, and certification rules?
- How can deployment feedback improve learning and synthesis without invalidating runtime safety guarantees?

IV. Towards Generative AI in the Physical World

Our proposed uncertainty-aware pipeline can also support the safe deployment of generative AI (GenAI) in physical and human-in-the-loop systems. In these settings, GenAI models may generate plans [54], [55], instructions [56], or control-relevant outputs [57], [58] that influence real-world actions. Safety therefore depends not only on the quality of the generated content, but also on whether the system can reason about hidden environment states, uncertain user intent, and deployment-time distribution shift.ß Robust POMDPs provide a comprehensive way to extend our framework to GenAI-enabled physical AI. The hidden state can capture the physical environment, human intent, task context, and latent safety risks, while observations can include sensor data, prompts, dialogue history, and human feedback. Robust synthesis improves reliability under uncertainty in user behaviour, sensing, environment dynamics, and deployment conditions.

Runtime certification then acts as a safeguard by checking whether a generated output is safe before it affects the physical system, preventing uncertified GenAI outputs from causing unsafe behaviour. In this sense, the proposed uncertainty-aware pipeline offers a principled foundation for deploying GenAI in safety-critical physical systems.

V. Conclusion

In this paper, we have argued that robust POMDPs offer a principled foundation for Safe Physical AI under hidden state, noisy sensing, uncertain dynamics, and deployment-time distribution shift. Within this framework, robust offline synthesis aims to produce policies that generalise across plausible deployment models, while runtime certification provides an additional safeguard by checking proposed actions before execution. We have also discussed how this train-certify perspective can be extended to GenAI-enabled physical systems, where generated plans or instructions may influence real-world behaviour.

Future work should address four key directions: scalable synthesis for robust constrained POMDPs, data-driven construction of uncertainty sets, tractable runtime certificates for realistic CPS, and counterexample-guided refinement methods that maintain safety guarantees during learning. Progress on these directions will be necessary for building physical AI systems that can operate reliably under incomplete information and changing real-world conditions.

REFERENCES

[1] R. Baheti and H. Gill, "Cyber-physical systems," *The Impact of Control Technology*, vol. 12, no. 1, pp. 161–166, 2011.

[2] P. Radanliev, D. De Roure, M. Van Kleek, O. Santos, and U. Ani, "Artificial intelligence in cyber physical systems," *AI & Society*, vol. 36, no. 3, pp. 783–796, 2021.

[3] Z. Lv, D. Chen, R. Lou, and A. Alazab, "Artificial intelligence for securing industrial-based cyber–physical systems," *Future Generation Computer Systems*, vol. 117, pp. 291–298, 2021.

[4] L. Sha, S. Gopalakrishnan, X. Liu, and Q. Wang, "Cyber-physical systems: A new frontier," in *2008 IEEE international conference on sensor networks, ubiquitous, and trustworthy computing (sutc 2008)*. IEEE, 2008, pp. 1–9.

[5] S. Mitra, *Verifying Cyber-Physical Systems: A Path to Safe Autonomy*. MIT Press, 2021.

[6] B. A. Salau, A. Rawal, and D. B. Rawat, "Recent advances in artificial intelligence for wireless internet of things and cyber–physical systems: A comprehensive survey," *IEEE Internet of Things Journal*, vol. 9, no. 15, pp. 12 916–12 930, 2022.

[7] S. Gu, L. Yang, Y. Du, G. Chen, F. Walter, J. Wang, and A. Knoll, "A review of safe reinforcement learning: Methods, theories, and applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 12, pp. 11 216–11 235, 2024.

[8] N. Nikolakis, V. Maratos, and S. Makris, "A cyber physical system (cps) approach for safe human-robot collaboration in a shared workplace," *Robotics and Computer-Integrated Manufacturing*, vol. 56, pp. 233–243, 2019.

[9] W. Wiesemann, D. Kuhn, and B. Rustem, "Robust Markov decision processes," *Mathematics of Operations Research*, vol. 38, no. 1, pp. 153–183, 2013.

[10] M. L. Puterman, *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.

[11] T. Osogami, "Robust partially observable Markov decision process," in *Proceedings of the International Conference on Machine Learning*. PMLR, 2015, pp. 106–115.

[12] M. Cubuktepe, N. Jansen, S. Junges, A. Marandi, M. Suilen, and U. Topcu, "Robust finite-state controllers for uncertain POMDPs," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 13, 2021, pp. 11 792–11 800.

[13] E. M. Bovy, M. Suilen, S. Junges, and N. Jansen, "Imprecise probabilities meet partial observability: Game semantics for robust POMDPs," *arXiv preprint arXiv:2405.04941*, 2024.

[14] M. Krale, E. M. Bovy, M. F. Galesloot, T. Simão, and N. Jansen, "On evaluating policies for robust POMDPs," *Proceedings of the Advances in Neural Information Processing Systems*, vol. 38, pp. 57 233–57 272, 2026.

[15] K. Chatterjee, M. Chmelik, and M. Tracol, "What is decidable about partially observable Markov decision processes with ω-regular objectives," *Journal of Computer and System Sciences*, vol. 82, no. 5, pp. 878–911, 2016.

[16] H. Xu and S. Mannor, "Distributionally robust Markov decision processes," *Proceedings of the Advances in Neural Information Processing Systems*, vol. 23, 2010.

[17] H. Zhang, H. Chen, C. Xiao, B. Li, M. Liu, D. Boning, and C.-J. Hsieh, "Robust deep reinforcement learning against adversarial perturbations on state observations," *Proceedings of the Advances in Neural Information Processing Systems*, vol. 33, pp. 21 024–21 037, 2020.

[18] H. Itoh and K. Nakamura, "Partially observable Markov decision processes with imprecise parameters," *Artificial Intelligence*, vol. 171, no. 8-9, pp. 453–490, 2007.

[19] Y. Ni and Z.-Q. Liu, "Bounded-parameter partially observable Markov decision processes." in *ICAPS*, 2008, pp. 240–247.

[20] M. F. Galesloot, M. Suilen, T. D. Simão, S. Carr, M. T. Spaan, U. Topcu, and N. Jansen, "Pessimistic iterative planning with RNNs for robust POMDPs," *arXiv preprint arXiv:2408.08770*, 2024.

[21] A. Undurti and J. P. How, "An online algorithm for constrained POMDPs," in *Proceedings of the IEEE International Conference on Robotics and Automation*. IEEE, 2010, pp. 3966–3973.

[22] M. Khonji, A. Jasour, and B. C. Williams, "Approximability of constant-horizon constrained POMDP," in *Proceedings of the International Joint Conference on Artificial Intelligence*, 2019, pp. 5583–5590.

[23] J. Lee, G.-H. Kim, P. Poupart, and K.-E. Kim, "Monte-Carlo tree search for constrained POMDPs," *Proceedings of Advances in Neural Information Processing Systems*, vol. 31, 2018.

[24] S. Carr, N. Jansen, S. Junges, and U. Topcu, "Safe reinforcement learning via shielding for POMDPs," *arXiv preprint arXiv:2204.00755*, 2022.

[25] W. Scarbro and R. Mangal, "Interval pomdp shielding for imperfect-perception agents," *arXiv preprint arXiv:2604.20728*, 2026.

[26] O. Madani, S. Hanks, and A. Condon, "On the undecidability of probabilistic planning and infinite-horizon partially observable markov decision problems," *Aaai/iaai*, vol. 10, no. 315149.315395, 1999.

[27] A. Bork, J.-P. Katoen, and T. Quatmann, "Under-approximating expected total rewards in pomdps," in *International Conference on Tools and Algorithms for the Construction and Analysis of Systems*. Springer, 2022, pp. 22–40.

[28] W. Scarbro, C. Imrie, S. G. Yaman, K. Fatehi, C. Păsăreanu, R. Calinescu, and R. Mangal, "Conformal safety shielding for imperfect-perception agents," in *Proceedings of the International Conference on Runtime Verification*. Springer, 2025, pp. 213–232.

[29] C. Zhou, Y. Gao, and P. Yu, "Ensuring logic in the fog: Sound POMDP synthesis with LTL objectives," in *Proceedings of the 35th International Joint Conference on Artificial Intelligence*, 2026. [Online]. Available: https://arxiv.org/abs/2605.12581

[30] E. Altman, *Constrained Markov decision processes*. Routledge, 2021.

[31] A. Wachi and Y. Sui, "Safe reinforcement learning in constrained Markov decision processes," in *Proceedings of the International Conference on Machine Learning*. PMLR, 2020, pp. 9797–9806.

[32] R. Singh, A. Gupta, and N. B. Shroff, "Learning in constrained Markov decision processes," *IEEE Transactions on Control of Network Systems*, vol. 10, no. 1, pp. 441–453, 2022.

[33] V. Borkar and R. Jain, "Risk-constrained Markov decision processes," *IEEE Transactions on Automatic Control*, vol. 59, no. 9, pp. 2574–2579, 2014.

[34] D.-S. Zois, M. Levorato, and U. Mitra, "Active classification for POMDPs: A Kalman-like state estimator," *IEEE Transactions on Signal Processing*, vol. 62, no. 23, pp. 6209–6224, 2014.

[35] S. Sheng, P. Yu, D. Parker, M. Kwiatkowska, and L. Feng, "Safe POMDP online planning among dynamic agents via adaptive conformal prediction," *IEEE Robotics and Automation Letters*, vol. 9, no. 11, pp. 9946–9953, 2024.

[36] S. Junges, N. Jansen, and S. A. Seshia, "Enforcing almost-sure reachability in POMDPs," in *Proceedings of the International Conference on Computer Aided Verification*. Springer, 2021, pp. 602–625.

[37] M. Ahmadi, N. Jansen, B. Wu, and U. Topcu, "Control theory meets POMDPs: A hybrid systems approach," *IEEE Transactions on Automatic Control*, vol. 66, no. 11, pp. 5191–5204, 2020.

[38] M. Ahmadi, A. Singletary, J. W. Burdick, and A. D. Ames, "Safe policy synthesis in multi-agent POMDPs via discrete-time barrier functions," in *Proceedings of the IEEE 58th Conference on Decision and Control (CDC)*. IEEE, 2019, pp. 4797–4803.

[39] J. Zhou, L. Yan, J. Liang, and K. Yang, "Safe reinforcement learning for mixed-autonomy platoons: A cooperative control barrier function approach," in *Proceedings of the American Control Conference*. IEEE, 2025, pp. 361–367.

[40] Y. Gao, K. H. Johansson, and L. Xie, "Computing probabilistic controlled invariant sets," *IEEE Transactions on Automatic Control*, vol. 66, no. 7, pp. 3138–3151, 2020.

[41] Y. Gao, M. Cannon, L. Xie, and K. H. Johansson, "Invariant cover: Existence, cardinality bounds, and computation," *Automatica*, vol. 129, 2021.

[42] Z. Liu, H. Chen, Y. Gao, and N. Ozay, "Opportunistic safety outside the maximal controlled invariant set," *IEEE Control Systems Letters*, vol. 7, pp. 3992–3997, 2023.

[43] Y. Wang, S. Chaudhuri, and L. E. Kavraki, "Bounded policy synthesis for POMDPs with safe-reachability objectives," *arXiv preprint arXiv:1801.09780*, 2018.

[44] Y. Wang, A. A. R. Newaz, J. D. Hernández, S. Chaudhuri, and L. E. Kavraki, "Online partial conditional plan synthesis for POMDPs with safe-reachability objectives: Methods and experiments," *IEEE Transactions on Automation Science and Engineering*, vol. 18, no. 3, pp. 932–945, 2021.

[45] M. Vahs, J. Verhagen, and J. Tumova, "Safety-critical control under partial observability: Reach-avoid POMDP meets belief space control," *arXiv preprint arXiv:2603.10572*, 2026.

[46] Y. Gao, A. Abate, L. Xie, and K. H. Johansson, "Distributional reachability for markov decision processes: Theory and applications," *IEEE Transactions on Automatic Control*, vol. 69, no. 7, pp. 4598–4613, 2023.

[47] Y. Gao, P. Yu, D. V. Dimarogonas, K. H. Johansson, and L. Xie, "Robust self-triggered control for time-varying and uncertain constrained systems via reachability analysis," *Automatica*, vol. 107, pp. 574–581, 2019.

[48] N. Jansen, B. Könighofer, S. Junges, A. Serban, and R. Bloem, "Safe reinforcement learning using probabilistic shields," in *Proceedings of the 31st International Conference on Concurrency Theory*. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2020, pp. 3–1.

[49] F. Belardinelli, A. W. Goodall *et al.*, "Probabilistic shielding for safe reinforcement learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, no. 15, 2025, pp. 16 091–16 099.

[50] V. Venktesh, M. Rathee, and A. Anand, "Trust but verify! a survey on verification design for test-time scaling," *arXiv preprint arXiv:2508.16665*, 2025.

[51] T. A. Henzinger, R. Jhala, and R. Majumdar, "Counterexample-guided control," in *International Colloquium on Automata, Languages, and Programming*. Springer, 2003, pp. 886–902.

[52] J. Seipp and M. Helmert, "Counterexample-guided cartesian abstraction refinement for classical planning," *Journal of Artificial Intelligence Research*, vol. 62, pp. 535–577, 2018.

[53] B. Gangopadhyay and P. Dasgupta, "Counterexample guided rl policy refinement using bayesian optimization," *Advances in Neural Information Processing Systems*, vol. 34, pp. 22 783–22 794, 2021.

[54] G. Dagan, F. Keller, and A. Lascarides, "Dynamic planning with an LLM," *arXiv preprint arXiv:2308.06391*, 2023.

[55] X. Huang, W. Liu, X. Chen, X. Wang, H. Wang, D. Lian, Y. Wang, R. Tang, and E. Chen, "Understanding the planning of LLM agents: A survey," *arXiv preprint arXiv:2402.02716*, 2024.

[56] Y. Wang, Z. Yu, W. Yao, Z. Zeng, L. Yang, C. Wang, H. Chen, C. Jiang, R. Xie, J. Wang *et al.*, "Pandalm: An automatic evaluation benchmark for llm instruction tuning optimization," in *Proceedings of the International Conference on Learning Representations*, vol. 2024, 2024, pp. 43 573–43 593.

[57] T.-H. Wu, L. Lian, J. E. Gonzalez, B. Li, and T. Darrell, "Self-correcting LLM-controlled diffusion models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 6327–6336.

[58] S. Zhao, D. Chen, Y.-C. Chen, J. Bao, S. Hao, L. Yuan, and K.-Y. K. Wong, "Uni-controlnet: All-in-one control to text-to-image diffusion models," *Proceedings of the Advances in Neural Information Processing Systems*, vol. 36, pp. 11 127–11 150, 2023.