# Reinforcement Learning Enabled Design of Electronic-State Engineering Strategies for Sustainable Oxygen Evolution Catalysis

Kekang Duan
Department of Computer Science and Engineering, University at Buffalo, Buffalo, NY, USA.
duan24@buffalo.edu

Jack Battler
Department of Computer Science, University of Central Florida, Orlando, FL, USA.
jan967@ucf.edu

Bastian Ceaster
School of Computing, Clemson University, Clemson, SC, USA.
contactbastian@clemson.edu

Erthur Wignir
Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, MO, USA.
contactarthur@missouri.edu

**Abstract**

The oxygen evolution reaction (OER) constitutes a critical bottleneck in sustainable energy conversion systems, yet the design of high-performance OER catalysts remains constrained by the vast and combinatorially complex landscape of electronic states that govern catalytic activity. Traditional approaches to electronic-state engineering rely on trial-and-error experimentation or density functional theory calculations that are ill-suited to navigate high-dimensional design spaces under operational constraints. This paper presents a systems-level framework for leveraging reinforcement learning (RL) to autonomously discover and optimize electronic-state engineering strategies for transition metal oxide catalysts. We conceptualize the catalyst design problem as a sequential decision-making process in which an RL agent modulates synthesis parameters, compositional levers, and defect chemistries to tune frontier electronic states such as metal-oxygen covalency, charge-transfer excitations, and correlated electron configurations. The architecture integrates a multi-fidelity simulation environment, active experimental validation loops, and a distributed computing infrastructure that balances exploration-exploitation trade-offs across heterogeneous computational resources. We provide a detailed examination of structural trade-offs in representing electronic state spaces, formulating reward signals that reconcile catalytic activity with long-term stability, and governing the socio-technical deployment of autonomous discovery platforms. Robustness considerations under epistemic uncertainty from both computational approximations and experimental variabilities are analyzed through the lens of domain randomization and ensemble policy methods. Furthermore, we address fairness and policy implications by proposing open-access benchmarking ecosystems and distributed manufacturing models that ensure globally equitable access to sustainably designed catalytic materials. This work reframes catalyst discovery as a large-scale adaptive system, offering

governance principles and infrastructure blueprints that extend beyond OER to broader materials design challenges in the transition to a carbon-neutral energy infrastructure.

**Keywords**

reinforcement learning, oxygen evolution reaction, electronic-state engineering, sustainable catalysis, autonomous materials discovery, systems architecture.

## 1. Introduction

The global transition toward sustainable energy systems hinges on the efficient production of clean fuels through electrochemical water splitting, wherein the oxygen evolution reaction (OER) represents the kinetic bottleneck that limits overall energy efficiency [1]. Transition metal oxides have emerged as promising OER catalysts owing to their earth abundance and tunable electronic properties, yet the precise engineering of their electronic states to achieve optimal activity and durability remains an open challenge [2]. The catalytic performance of these materials is governed by a delicate interplay of orbital hybridization, charge-transfer energies, spin configurations, and electron correlation effects that collectively define the adsorption energetics of reaction intermediates [3]. Navigating this multidimensional electronic structure space to identify robust catalyst formulations demands a paradigm shift away from traditional Edisonian trial-and-error methodologies and toward systematic, autonomous design frameworks.

Conventional catalyst design pipelines rely heavily on density functional theory (DFT) calculations to screen hypothetical structures, but the computational cost of high-throughput DFT and its intrinsic approximations in treating strongly correlated oxides limit the pace and reliability of discovery [4]. Moreover, experimental synthesis and characterization workflows introduce additional layers of complexity stemming from non-equilibrium phase formation, defect distributions, and surface reconstructions under operando conditions. These factors render the electronic state engineering problem ill-posed for gradient-based optimization or exhaustive enumeration. Reinforcement learning, a branch of machine learning concerned with learning optimal policies through interaction with an environment, offers a principled approach to sequential decision-making under uncertainty [5]. By framing catalyst design as a Markov decision process, RL agents can learn to propose synthesis protocols and compositional adjustments that traverse the electronic state landscape while adaptively allocating resources between exploration of novel configurations and exploitation of known high-performance regions.

The present work develops a comprehensive systems architecture for RL-enabled electronic-state engineering in OER catalysis. Rather than focusing narrowly on algorithmic novelty, we adopt a socio-technical perspective that integrates computational modeling, experimental automation, data infrastructure, and governance mechanisms into a coherent adaptive system. The paper makes three primary contributions. First, we articulate a modular system architecture that decouples state representation, policy learning, and reward computation while enabling seamless integration of multi-fidelity data sources from DFT, machine-learned interatomic potentials, and robotic experimentation. Second, we analyze structural trade-offs inherent in designing electronic-state descriptors, reward shaping that balances activity metrics with stability and synthesizability constraints, and the management of epistemic uncertainty through robust RL techniques. Third, we extend the discussion beyond mere technical optimization to address fairness, accessibility, and policy dimensions, including how autonomous discovery platforms can be governed to prevent the concentration of catalytic

knowledge and manufacturing capacity within a narrow set of industrialized nations. This systems-oriented inquiry is intended to guide the design of next-generation autonomous laboratories that accelerate the discovery of sustainable catalytic materials while aligning with broader societal goals.

## 2. Background and Related Work

The relationship between electronic structure and OER activity has been extensively studied in the context of descriptor-based approaches. The occupancy of the transition metal eg orbital, the position of the O p-band center relative to the metal d-band, and the degree of metal-oxygen covalency have all been identified as key determinants of the binding strength of oxygenated intermediates [6]. In strongly correlated oxides, phenomena such as ligand-hole formation and charge-transfer excitations further enrich the design space, as exemplified by the two-step oxidation pathway that forms a Zhang-Rice singlet state to trigger water oxidation under operando conditions [7]. Such discoveries underscore the fact that optimal electronic configurations often arise from non-intuitive cooperative effects that cannot be anticipated from single-parameter descriptors. Consequently, the research community has increasingly turned to data-driven methods, including high-throughput DFT and machine learning surrogate models, to accelerate screening. Yet these methods typically operate in a batch mode, treating each candidate material as an independent evaluation, which forgoes the opportunity to learn from the trajectory of sequential experiments.

Early efforts to apply RL in materials science have demonstrated success in domains such as crystal structure prediction, molecular design, and process optimization [8]. In the context of heterogeneous catalysis, recent works have used RL to optimize nanoparticle morphologies, alloy compositions, or reaction conditions. However, the specific challenge of electronic-state engineering for oxide catalysts remains underexplored because electronic configuration is an emergent property that depends on synthesis history, defect chemistry, and external stimuli in a path-dependent manner. This path dependence aligns naturally with the sequential nature of RL, where an agent's actions alter the system state and influence future possibilities. Prior research on active learning loops for catalysis has shown that iterative experimental design can substantially reduce the number of experiments required to identify optimal catalysts, but these approaches typically employ myopic acquisition functions rather than the temporally extended credit assignment of RL [9]. By adopting a full RL formulation, one can optimize over entire trajectories of synthesis, post-treatment, and electrochemical conditioning, capturing long-term trade-offs that are invisible to one-step look-ahead strategies.

The integration of RL with automated experimentation raises significant infrastructure challenges that have received limited attention in the literature. High-throughput experimental platforms, such as combinatorial sputtering systems and scanning droplet cells, can generate large volumes of composition-structure-activity data, but closing the loop with RL requires robust middleware for data streaming, model retraining, and uncertainty quantification [10]. Similarly, computational environments must manage multiple fidelity levels, ranging from semi-empirical electronic structure methods to hybrid DFT and dynamical mean-field theory, each imposing different latency and accuracy profiles. The design of the RL agent's observation and action spaces must therefore reconcile the semantic gap between quantum-mechanical descriptors and laboratory-accessible control variables such as precursor concentrations, annealing temperatures, and electrochemical cycling protocols. This interface challenge constitutes a central theme of the present work.

## 3. System Architecture for RL-Driven Electronic-State Engineering

We propose a layered system architecture comprising four interconnected subsystems: the environment interface layer, the state representation and featurization layer, the policy learning core, and the orchestration and governance layer. The environment interface layer abstracts the physical and computational resources available to the agent. On the computational side, this includes a hierarchy of models ranging from low-fidelity bond-valence sum calculations to high-fidelity DFT plus U and hybrid functional simulations, as well as machine-learned force fields trained on ab initio data for accelerated structural relaxation. On the experimental side, the layer interfaces with automated synthesis robots and high-throughput electrochemical testing stations through standardized communication protocols. Crucially, the interface layer incorporates a multi-fidelity scheduler that dynamically allocates queries to simulators or laboratory instruments based on the expected value of information and the current computational budget, embodying a resource-aware decision-making framework that is essential for sustainable and cost-effective operation [11].

The state representation layer transforms raw measurement and simulation outputs into a structured observation vector that captures the relevant electronic-state features of a catalyst candidate. Designing this representation involves navigating a fundamental trade-off between completeness and generalizability. High-dimensional representations, such as the full density of states projected onto atomic orbitals, offer maximal information but suffer from the curse of dimensionality and poor transferability across different oxide families. Low-dimensional descriptors, such as the eg filling, charge-transfer energy, and oxygen 2p band center, facilitate efficient learning but risk excluding emergent electronic phenomena such as polaron formation or dynamic Jahn-Teller distortions. We advocate a hierarchical representation scheme in which a graph neural network encoder maps the local atomic environment and oxidation states to a latent space that is refined through contrastive learning on experimental activity labels [12]. This latent representation serves as the input to the RL policy, enabling the agent to reason about electronic structure without being tied to a single fixed descriptor set. The encoder is periodically updated as new experimental data become available, ensuring that the representation co-evolves with the expanding knowledge base.

The policy learning core implements a suite of RL algorithms, with a primary emphasis on actor-critic architectures that maintain separate networks for state-value estimation and action selection. Given the continuous nature of many synthesis parameters—such as temperature, partial pressure of reactive gases, and precursor molar ratios—we employ deterministic policy gradient methods with hindsight experience replay to handle sparse reward signals that are common in catalysis when a substantial improvement is only observed after multiple sequential treatments [13]. The action space is structured hierarchically to reflect the natural decomposition of catalyst synthesis into stages: precursor selection, synthesis condition setting, post-synthetic treatment, and electrochemical conditioning protocol. A meta-controller learns to select among these stages, while low-level controllers optimize continuous parameters within each stage. This hierarchical decomposition not only improves sample efficiency but also aligns with the organizational logic of human experimentalists, facilitating interpretability and human-in-the-loop oversight.

The orchestration and governance layer manages the interaction between the agent, computational infrastructure, and laboratory automation, while also enforcing safety constraints and ethical guidelines. It implements a distributed message-passing framework based on the actor model of concurrency, enabling the system to scale horizontally across multiple compute clusters and geographically dispersed experimental facilities [14]. Within

this layer, a policy governance module monitors the distribution of proposed experiments to ensure that the agent does not persistently allocate resources to materials that would require critical raw materials sourced from conflict-affected regions or that would pose disproportionate environmental hazards during synthesis. Such governance mechanisms transform the RL system from a purely optimization-driven engine into a socio-technical infrastructure that can be aligned with sustainability principles beyond mere catalytic efficiency.

## 4. Design Space Representation and Reward Formulation

The representation of electronic states for RL-driven catalyst design demands a careful articulation of the state variables that are both causally linked to catalytic function and controllable in practice. At the electronic level, key quantities include the metal d-band center relative to the Fermi level, the ligand field splitting energy, the hybridization matrix elements between the metal d and oxygen p states, and the Hubbard U parameter that quantifies on-site Coulomb repulsion. These quantities are not directly tunable but are functions of the local coordination environment, oxidation state, and cation-anion stoichiometry. The RL agent must therefore act on the atomic-scale configuration—cation substitution, vacancy creation, strain engineering—and observe the resulting electronic structure through a combination of spectroscopic signatures and computed electronic properties. This mapping from configurational actions to electronic states is highly nonlinear and often exhibits hysteresis due to irreversible structural transformations, making it a challenging partially observable Markov decision process in which the full electronic details are never directly accessible to the agent but must be inferred from noisy experimental and computational probes [15].

To address partial observability, we incorporate recurrent neural network layers within the policy and value networks, allowing the agent to maintain an internal belief state over the catalyst's true electronic configuration. This belief state integrates information across sequential synthesis and characterization steps, enabling the agent to disambiguate between catalysts that appear similar based on instantaneous observations but have distinct hidden properties such as defect ordering or metastable phase fractions. The recurrent architecture also provides a natural mechanism for the agent to learn temporal patterns, such as the kinetics of oxygen vacancy diffusion during annealing or the evolution of surface reconstruction layers during electrochemical cycling. The inclusion of memory transforms the RL system from a reactive controller into a deliberative planner that can orchestrate multi-step treatments to steer a catalyst into a desired electronic state basin.

Reward formulation is a critical design choice that embodies the values and objectives of the catalytic system. The most straightforward reward signal is the overpotential at a given current density, which directly quantifies the thermodynamic efficiency loss during OER. However, optimizing solely for low overpotential can lead to catalysts that are highly active initially but degrade rapidly due to cation leaching or amorphization. Therefore, the reward function must incorporate a durability component, such as the retention of activity over a specified number of accelerated stress test cycles, as well as a measure of faradaic efficiency to penalize parasitic side reactions. Multi-objective RL frameworks allow the agent to learn a Pareto front of policies that balance activity, stability, and other metrics, with the governance layer subsequently selecting among these policies based on deployment context [16]. Additional soft penalties can discourage the use of scarce or toxic elements, thereby embedding sustainability constraints directly into the optimization objective rather than relegating them to post hoc screening.

The sparsity of feedback signals in catalysis poses a significant challenge for credit assignment. Many synthesis actions precede the final performance evaluation by several days of experimental processing, and the relationship between an early precursor choice and the ultimate activity may be obscured by subsequent thermal treatments. We mitigate this challenge through a combination of auxiliary prediction tasks, such as predicting intermediate spectroscopic observables from action sequences, and by shaping the reward with physics-informed metrics computed from rapid DFT evaluations [17]. These intermediate rewards serve as waypoints that guide the agent's exploration toward regions of the design space where electronic states are more likely to yield favorable OER kinetics, substantially accelerating learning in early phases before a sufficient volume of terminal experimental rewards has accumulated.

## 5. Infrastructure and Scalability Considerations

The deployment of RL-enabled electronic-state engineering at scale requires a robust computational and experimental infrastructure that can sustain high-throughput autonomous operation over extended periods. On the computational side, the primary bottleneck lies in the generation of training data for the encoder and the reward computation via DFT. While single-point DFT calculations on oxide surfaces have become routine, the broader exploration undertaken by the RL agent frequently requires relaxation of defected supercells and nudged elastic band calculations for reaction barriers, which scale superlinearly with system size. To manage this load, we adopt an elastic cloud computing strategy that dynamically provisions high-performance computing nodes in response to the agent's current exploration demands, combined with a priority queuing system that schedules urgent calculations needed for ongoing policy rollouts ahead of background simulation campaigns [18]. This elastic infrastructure prevents resource contention and enables the agent to maintain a consistent data ingestion rate, which is critical for stable learning dynamics in online RL settings.

Experimental automation presents a distinct set of infrastructure challenges related to sample handling, reproducibility, and instrument drift. We envision a modular robotic laboratory in which synthesis stations, thermal processing units, and electrochemical characterization cells are integrated through a centralized laboratory information management system. Each physical action proposed by the agent, such as dispensing a specific volume of precursor solution or setting a furnace ramp profile, is translated into a sequence of instrument commands via a translation layer that enforces hardware constraints and safety interlocks. Crucially, the system must incorporate automated calibration protocols and periodic quality control measurements to detect and correct for sensor drift, as systematic errors in the observed state can lead to policy degradation when the agent is no longer learning in the same environment it was trained for [19]. The infrastructure must therefore be self-monitoring, with anomaly detection algorithms that trigger maintenance workflows when deviations exceed pre-defined thresholds.

The distributed nature of the system, potentially spanning multiple laboratories across different time zones, introduces additional governance requirements related to data provenance, experiment provenance, and model versioning. Each experimental datum must be accompanied by metadata that captures the full provenance chain, including the identity of the instruments, their calibration status, the operator actions, and the software versions of the agent and analysis pipelines. This provenance tracking is essential not only for scientific reproducibility but also for the accountability of autonomous decision-making in the event of unsafe or erroneous proposals. We advocate the adoption of cryptographically signed

provenance logs based on distributed ledger technology to ensure immutability and transparency across collaborating institutions [20]. Such infrastructure investments, while substantial, are a prerequisite for building trust in autonomous materials discovery systems and for enabling the regulatory acceptance that will be necessary as these technologies mature and begin to influence commercial manufacturing.

## 6. Robustness, Fairness, and Policy Implications

The deployment of RL for catalyst design occurs within a socio-technical context where the distribution of benefits and risks is shaped by the design choices embedded in the system. Robustness under epistemic uncertainty is a primary technical concern, as the models that the agent relies upon—DFT with approximate exchange-correlation functionals, machine-learned interatomic potentials, and surrogate models for electrochemical stability—all possess systematic errors that can lead the policy to overestimate the performance of certain material families. Domain randomization, in which the agent is trained across an ensemble of computational models that span plausible physical parameter ranges, has proven effective in transferring policies from simulation to reality in robotics, and we adapt this principle to the catalysis domain by randomizing Hubbard U values, strain states, and surface terminations during training [21]. The resulting policies exhibit greater robustness when transferred to experimental settings where the exact electronic structure deviates from any single computational prediction. Furthermore, ensemble policy methods that aggregate the recommendations of multiple agents trained under different model assumptions provide a mechanism for quantifying decision uncertainty and for implementing risk-averse action selection when the cost of a failed experiment is high.

Beyond robustness, fairness considerations demand attention in the design of autonomous catalyst discovery platforms. High-throughput experimental facilities and the computational resources required for RL-driven optimization are capital-intensive, which creates a risk that the resulting catalytic knowledge and intellectual property will be concentrated in a small number of well-funded institutions, predominantly in high-income countries. This concentration could perpetuate or even exacerbate existing inequities in access to sustainable energy technologies, as patent-protected catalyst formulations may be licensed under terms that are prohibitive for manufacturers in developing economies. To countervail this tendency, we advocate for the establishment of open-access catalyst design ecosystems that curate publicly available datasets, pre-trained models, and benchmarking suites, analogous to the role that open-source software foundations have played in democratizing access to artificial intelligence tools [22]. Governance frameworks should include licensing models that facilitate royalty-free use of autonomously discovered catalyst compositions for applications in regions classified as low- and middle-income by multilateral development agencies, while still permitting commercial exploitation in high-income markets to incentivize investment.

The policy implications of RL-enabled electronic-state engineering extend to energy diplomacy, supply chain resilience, and environmental regulation. As autonomous discovery platforms accelerate the pace of catalyst innovation, regulatory agencies will need to develop new capacity for risk assessment of computationally designed materials that may not have close analogues among substances with established safety profiles. Precautionary frameworks that require life-cycle assessment and toxicity screening as integral components of the reward function, rather than as downstream gatekeeping steps, can embed environmental stewardship into the optimization process from the outset [23]. At the geopolitical level, the ability to rapidly design catalysts that avoid reliance on critical minerals—such as iridium, which is

currently essential for proton-exchange membrane electrolyzers but is among the rarest elements in the Earth's crust—has profound implications for energy security and international relations. RL-driven substitution of scarce elements with abundant alternatives, guided by electronic-state engineering principles that replicate catalytic function through alternative orbital hybridization schemes, represents a strategic capability that nations will seek to cultivate as part of their clean energy transition portfolios.

The governance of autonomous discovery platforms must also contend with labor market transformations in the scientific workforce. As RL agents take on an increasing share of experimental design and data interpretation tasks, the role of human researchers will shift toward system oversight, creative hypothesis generation regarding novel electronic mechanisms, and the articulation of societal values that inform reward design. Training programs that equip materials scientists and chemists with the computational literacy to engage critically with RL-driven systems, and that foster interdisciplinary collaboration with computer scientists and ethicists, will be essential for ensuring that the human capital pipeline remains aligned with the evolving demands of the field [24]. Integrating human feedback into the RL loop through preference-based reward learning offers a pathway for embedding practitioner expertise and societal norms into the objective function without requiring explicit specification of all constraints, thereby creating a more responsive and democratically legitimate optimization process.

## 7. Conclusion

This paper has presented a systems-level framework for the application of reinforcement learning to electronic-state engineering in sustainable oxygen evolution catalysis. By treating catalyst design as a sequential decision-making problem under uncertainty, the RL paradigm enables the autonomous navigation of high-dimensional electronic configuration spaces that are intractable for conventional screening methods. We have articulated a modular architecture that spans multi-fidelity simulation, automated experimentation, and governance mechanisms, and we have examined the structural trade-offs inherent in representing electronic states, formulating multi-objective rewards, and managing infrastructure at scale. The analysis of robustness, fairness, and policy implications situates the technical system within its broader socio-technical context, highlighting the imperative to design autonomous discovery platforms that not only accelerate innovation but also promote equitable access to sustainable energy technologies and respect planetary boundaries. As the global community accelerates toward net-zero emissions, the integration of reinforcement learning with materials science will play an increasingly critical role in unlocking the catalytic transformations that underpin a carbon-neutral energy economy, provided that the governance frameworks guiding these technologies evolve in tandem with their technical capabilities.

## References


1. Seh, Z. W., Kibsgaard, J., Dickens, C. F., Chorkendorff, I., Nørskov, J. K., & Jaramillo, T. F. (2017). Combining theory and experiment in electrocatalysis: Insights into materials design. Science, 355(6321), eaad4998.

2. Hwang, J., Rao, R. R., Giordano, L., Katayama, Y., Yu, Y., & Shao-Horn, Y. (2017). Perovskites in catalysis and electrocatalysis. Science, 358(6364), 751–756.

3. Suntivich, J., May, K. J., Gasteiger, H. A., Goodenough, J. B., & Shao-Horn, Y. (2011). A perovskite oxide optimized for oxygen evolution catalysis from molecular orbital principles. Science, 334(6061), 1383–1385.

4. Jain, A., Ong, S. P., Hautier, G., Chen, W., Richards, W. D., Dacek, S., Cholia, S., Gunter, D., Skinner, D., Ceder, G., & Persson, K. A. (2013). Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. APL Materials, 1(1), 011002.

5. Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction (2nd ed.). MIT Press.

6. Grimaud, A., Diaz-Morales, O., Han, B., Hong, W. T., Lee, Y. L., Giordano, L., Stoerzinger, K. A., Koper, M. T. M., & Shao-Horn, Y. (2017). Activating lattice oxygen redox reactions in metal oxides to catalyse oxygen evolution. Nature Chemistry, 9(5), 457–465.

7. Peng, C. K., Lin, Y. C., Chiang, C. L., Qian, Z., Huang, Y. C., Dong, C. L., ... & Lin, Y. G. (2023). Zhang-Rice singlets state formed by two-step oxidation for triggering water oxidation under operando conditions. Nature Communications, 14(1), 529.

8. Zhou, Z., Li, X., & Zare, R. N. (2017). Optimizing chemical reactions with deep reinforcement learning. ACS Central Science, 3(12), 1337–1344.

9. Lookman, T., Balachandran, P. V., Xue, D., & Yuan, R. (2019). Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design. npj Computational Materials, 5(1), 21.

10. Burger, B., Maffettone, P. M., Gusev, V. V., Aitchison, C. M., Bai, Y., Wang, X., Li, X., Alston, B. M., Li, B., Clowes, R., Rankin, N., Harris, B., Sprick, R. S., & Cooper, A. I. (2020). A mobile robotic chemist. Nature, 583(7815), 237–241.

11. Kandasamy, K., Dasarathy, G., Schneider, J., & Póczos, B. (2017). Multi-fidelity Bayesian optimisation with continuous approximations. Proceedings of the 34th International Conference on Machine Learning, 1799–1808.

12. Xie, T., & Grossman, J. C. (2018). Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. Physical Review Letters, 120(14), 145301.

13. Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., & Wierstra, D. (2016). Continuous control with deep reinforcement learning. Proceedings of the 4th International Conference on Learning Representations.

14. Moritz, P., Nishihara, R., Wang, S., Tumanov, A., Liaw, R., Liang, E., Elibol, M., Yang, Z., Paul, W., Jordan, M. I., & Stoica, I. (2018). Ray: A distributed framework for emerging AI applications. Proceedings of the 13th USENIX Symposium on Operating Systems Design and Implementation, 561–577.

15. Hausknecht, M., & Stone, P. (2015). Deep recurrent Q-learning for partially observable MDPs. Proceedings of the 2015 AAAI Fall Symposium Series.

16. Roijers, D. M., Vamplew, P., Whiteson, S., & Dazeley, R. (2013). A survey of multi-objective sequential decision-making. Journal of Artificial Intelligence Research, 48, 67–113.

17. Ng, A. Y., Harada, D., & Russell, S. (1999). Policy invariance under reward transformations: Theory and application to reward shaping. Proceedings of the 16th International Conference on Machine Learning, 278–287.

18. Xu, M., Hu, J., & Panda, D. K. (2021). Accelerating reinforcement learning on high-performance computing systems. Proceedings of the 2021 IEEE International Parallel and Distributed Processing Symposium, 677–686.

19. Chiang, H. T. L., Hsu, J., Fiser, M., Tapia, L., & Faust, A. (2019). RL-RRT: Kinodynamic motion planning via learning reachability estimators from RL policies. IEEE Robotics and Automation Letters, 4(4), 4299–4306.

20. Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3(1), 160018.

21. Tobin, J., Fong, R., Ray, A., Schneider, J., Zaremba, W., & Abbeel, P. (2017). Domain randomization for transferring deep neural networks from simulation to the real world. Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems, 23–30.

22. Stall, S., Yarmey, L., Cutcher-Gershenfeld, J., Hanson, B., Lehnert, K., Nosek, B., Parsons, M., Robinson, E., & Wyborn, L. (2019). Make scientific data FAIR. Nature, 570(7759), 27–29.

23. Zimmermann, J. B., Anastas, P. T., Erythropel, H. C., & Leitner, W. (2020). Designing for a green chemistry future. Science, 367(6476), 397–400.

24. National Academies of Sciences, Engineering, and Medicine. (2022). Automated research workflows for accelerated discovery: Closing the knowledge loop. The National Academies Press.