Hierarchical Reinforcement Learning for Dynamic Financial Risk Control

Elton Karlstrom
Department of Computer Science and Engineering, University of Nevada, Reno, Reno, NV, USA.
karlstrom983@unr.edu

Folker Misicher
Department of Electrical Engineering and Computer Science, University of Kansas, Lawrence, KS, USA.
misicher1980@ku.edu

Elton Lundstrom
School of Computing, Clemson University, Clemson, SC, USA.
contactelton@clemson.edu

Abstract

Dynamic financial risk control increasingly requires systems that can adapt across time scales, institutional boundaries, and market states while remaining auditable, fair, and resilient. Hierarchical reinforcement learning offers a structural response by separating strategic risk posture from tactical exposure management and operational execution. This paper examines hierarchical reinforcement learning as a system-level architecture for dynamic financial risk control, with attention to temporal abstraction, option discovery, multi-agent coordination, governance, deployment, and sustainability. It argues that the central challenge is not merely optimizing a policy in a simulated market but aligning layered decision rights with regulatory obligations, data provenance, model risk management, and human oversight. The analysis discusses structural trade-offs among adaptability, interpretability, latency, and capital efficiency. It considers regime shifts, market microstructure, liquidity constraints, and adversarial behavior as forces that destabilize flat reinforcement learning agents. It also explores how hierarchy can support fairness and contestability when upper layers encode policy constraints and lower layers execute bounded actions. The paper concludes that hierarchical reinforcement learning can improve dynamic risk control only when treated as a socio-technical infrastructure rather than a standalone algorithmic innovation. Successful deployment depends on robust simulation, continuous validation, clear escalation protocols, and governance mechanisms that make layered autonomy accountable across institutions.

Keywords

hierarchical reinforcement learning; financial risk control; socio-technical systems; model governance; regime shifts; multi-agent systems; fairness; operational resilience.

1. Introduction

Financial risk control has shifted from periodic, rule-bound oversight toward continuous, data-intensive, and increasingly automated decision systems. Traditional risk management frameworks emphasize capital adequacy, value at risk, stress testing, and limit enforcement, but they often struggle when market conditions change faster than governance cycles can

adapt [14], [15], [16]. Reinforcement learning has attracted attention because it can learn sequential policies from interaction, optimize long-horizon objectives, and respond to changing environments without relying solely on static rules [1]. Deep reinforcement learning has demonstrated that agents can learn complex behaviors from high-dimensional observations, which suggests potential for financial environments characterized by noisy data, delayed rewards, and adversarial participants [8]. However, applying flat reinforcement learning directly to financial risk control raises serious concerns about instability, opacity, regulatory defensibility, and operational safety.

Hierarchical reinforcement learning addresses some of these concerns by decomposing decision-making into layers of abstraction [2], [3]. Instead of a single policy mapping every market observation to every possible action, hierarchical architectures allow higher-level policies to set goals, constraints, or risk budgets while lower-level policies execute bounded tactical actions [4], [5]. This separation is attractive for financial risk control because risk decisions naturally occur at multiple time scales. A board or risk committee may set appetite quarterly, a treasury desk may adjust exposure daily, and an execution algorithm may manage orders in milliseconds. A hierarchical structure can mirror these institutional layers, creating a closer correspondence between algorithmic control and organizational authority.

The promise of hierarchy is not automatic. It introduces new coordination problems, credit assignment difficulties, and governance questions. If upper layers encode strategic intent but lower layers learn unsafe shortcuts, the system may fail in ways that are difficult to diagnose. If lower layers are tightly constrained, the system may lose the adaptability that motivated reinforcement learning in the first place. This paper therefore treats hierarchical reinforcement learning as a socio-technical architecture for dynamic financial risk control. It examines how temporal abstraction, multi-agent coordination, robustness under regime shifts, fairness, deployment infrastructure, and policy implications interact. The argument is that hierarchy can improve risk control only when it is designed as an accountable system of layered decision rights rather than as a purely algorithmic optimization stack.

2. Foundations of Hierarchical Reinforcement Learning

Hierarchical reinforcement learning builds on the broader framework of reinforcement learning, in which an agent learns to maximize cumulative reward through trial-and-error interaction with an environment [1]. The core difficulty in long-horizon tasks is that credit for success or failure may be delayed, sparse, or distributed across many actions. In financial risk control, this difficulty is pronounced because a decision to reduce leverage may prevent a loss that never appears, while a decision to increase exposure may generate profit before a hidden tail risk materializes. Hierarchical methods attempt to manage this complexity by introducing temporal abstraction, allowing the agent to reason with temporally extended actions rather than only primitive steps [3]. These extended actions, often called options or skills, can represent coherent behaviors such as hedging a currency exposure, reducing a concentrated position, or escalating a liquidity concern.

A central idea in hierarchical reinforcement learning is that higher-level policies can operate over abstract states and temporally extended actions, while lower-level policies handle the details of execution [2]. The MAXQ framework, for example, decomposes value functions across subtasks, enabling reuse and modular learning [4]. Other approaches use hierarchies of machines or options to structure control and improve exploration [5]. More recent work has explored learned hierarchies in which higher-level agents set latent goals and lower-level agents learn to achieve them [6]. Data-efficient hierarchical reinforcement learning further

shows that off-policy learning and goal relabeling can make hierarchy more practical in complex domains [7]. These advances matter for finance because risk control is not a single task but a family of related tasks: monitoring, hedging, rebalancing, reporting, escalation, and recovery.

However, hierarchy is not a cure for fundamental reinforcement learning challenges. Temporal abstraction can make credit assignment easier at one level while making it harder at another, because lower-level failures may be hidden by upper-level goal selection. Non-stationarity becomes more complex when multiple layers learn simultaneously, and reward shaping may create unintended incentives. In financial systems, these issues are amplified by low signal-to-noise ratios, heavy-tailed outcomes, and adversarial adaptation. A hierarchical agent that learns to suppress volatility in normal periods may accumulate hidden fragility that appears only during stress. Consequently, hierarchical reinforcement learning for financial risk control must be evaluated not only by return or loss reduction but by its ability to preserve safety margins, respect constraints, and explain its layered decisions to human overseers.

3. Dynamic Financial Risk Control as a Socio-Technical System

Dynamic financial risk control is not merely a prediction problem. It is a socio-technical system that combines data infrastructure, organizational roles, regulatory mandates, legal obligations, and market microstructure. Machine learning has been applied to portfolio construction, default prediction, fraud detection, and market forecasting [9], [10], [11]. These applications demonstrate value, but they also reveal that model performance depends heavily on data quality, feature stability, and the institutional context in which predictions are used. A risk model that performs well in backtests may fail when liquidity disappears, correlations shift, or counterparties change behavior. Risk management literature therefore emphasizes stress testing, scenario analysis, and capital buffers as complements to statistical models [14], [15]. Regulatory frameworks similarly require banks to hold capital against market risk and to validate internal models under changing conditions [16].

Financial markets are also characterized by regime shifts. Economic regimes can change abruptly, altering volatility, correlation, and the relationship between risk factors [12], [13]. A reinforcement learning agent trained in one regime may fail in another because its learned policy encodes assumptions that no longer hold. Hierarchy can help by assigning regime detection or risk posture adjustment to an upper layer while allowing lower layers to adapt execution within a narrower scope. Yet regime detection itself is uncertain, and misclassification can cause the system to shift posture at the wrong time. If the upper layer reacts too slowly, the system may remain exposed during a crisis. If it reacts too quickly, it may trigger procyclical behavior that amplifies market stress.

The socio-technical dimension also includes human accountability. Risk managers, compliance officers, auditors, and regulators must be able to understand why a system took a particular action, what constraints it was operating under, and who is responsible for the outcome. Hierarchical reinforcement learning can support this need by making the levels of decision-making explicit. For example, an upper layer might encode a risk appetite approved by a committee, while a lower layer executes within predefined limits. This separation can improve transparency if the interfaces between layers are documented and monitored. However, it can also obscure responsibility if the upper layer's goals are vague or if lower layers learn behaviors that are technically compliant but substantively risky. A recent perspective on trust-calibrated multi-agent decision intelligence under regime shifts

emphasizes that trust must be calibrated dynamically rather than assumed from model accuracy alone [17]. This insight applies directly to hierarchical risk systems, where trust in lower layers depends on the reliability of upper-layer goals and the observability of execution.

4. Architecture and Structural Trade-offs

The architecture of a hierarchical reinforcement learning system for financial risk control must balance several competing objectives. The first is temporal resolution. Upper layers may operate at quarterly, daily, or hourly intervals, while lower layers may operate at millisecond or second intervals. If the upper layer updates too frequently, it may overreact to noise and destabilize lower-level execution. If it updates too slowly, it may fail to adapt to changing market conditions. The second objective is action granularity. Upper layers may set risk budgets, exposure targets, or hedging goals, while lower layers choose specific instruments, order types, and timing. This division can reduce the combinatorial complexity of the action space, but it requires clear interfaces and consistent semantics.

A third trade-off concerns interpretability and adaptability. Flat reinforcement learning can be highly adaptive but difficult to interpret. Hierarchical architectures can improve interpretability by separating strategic intent from tactical execution, but they can also create emergent behaviors that are not obvious from either layer alone. If upper-layer goals are too abstract, lower layers may find unexpected ways to satisfy them. If goals are too prescriptive, the system becomes a rule-based automaton with limited learning capacity. This tension is familiar in reinforcement learning research on option discovery and goal generation [2], [6], [7]. In financial risk control, the balance must favor auditability because regulators and internal risk functions need to reconstruct decision pathways after adverse events.

A fourth trade-off involves centralization and decentralization. A centralized upper layer can enforce consistent risk policy across desks and business lines, but it may become a bottleneck and a single point of failure. Decentralized lower layers can respond quickly to local conditions, but they may drift from enterprise risk appetite. Multi-agent reinforcement learning offers tools for coordination, but it also introduces non-stationarity because each agent's learning changes the environment for others [23], [24]. In financial markets, this non-stationarity is not merely technical; it reflects strategic interaction among banks, funds, market makers, and regulators. A hierarchical system must therefore include coordination mechanisms such as shared constraints, communication protocols, and arbitration rules. Without them, layered autonomy can produce conflicting actions that increase rather than reduce risk.

5. Regime Shifts, Robustness, and Multi-Agent Coordination

Regime shifts are among the most difficult challenges for dynamic risk control. Statistical models often assume stable relationships, but financial markets exhibit structural breaks, volatility clustering, and changes in liquidity [12], [13]. A hierarchical reinforcement learning system can address regime shifts by separating regime inference from policy execution. An upper layer could maintain multiple risk models or scenario weights, while a lower layer adjusts execution based on the inferred regime. This design resembles regime-switching models in finance, but it adds learning and adaptation. The advantage is that the system can change posture without retraining every lower-level skill from scratch. The risk is that regime inference itself becomes a source of error, especially when regimes are rare and poorly labeled.

Robustness also requires attention to adversarial behavior. Financial markets include participants who may exploit predictable algorithms, manipulate prices, or create feedback loops. A reinforcement learning agent that learns to execute in a simulated environment may be vulnerable to real-world adversaries who detect and counteract its patterns. Research on concrete problems in artificial intelligence safety highlights concerns such as reward hacking, side effects, and distributional shift [22]. These concerns are directly relevant to financial risk control. A hierarchical agent might learn to satisfy a risk constraint in a narrow technical sense while increasing exposure to hidden tail risks. It might also exploit simulation artifacts, such as unrealistic liquidity assumptions, that do not hold in live markets.

Multi-agent coordination further complicates robustness. In a large financial institution, multiple desks and algorithms may operate simultaneously under a shared risk limit. If each desk uses its own reinforcement learning agent, the collective behavior can become unstable even if each agent is individually well-behaved. Hierarchical reinforcement learning can mitigate this by introducing a coordinating layer that allocates risk budgets and monitors aggregate exposure. However, the coordinating layer must contend with strategic behavior from lower layers, delayed information, and conflicting incentives. The literature on multi-agent reinforcement learning emphasizes the difficulty of convergence and equilibrium selection in such settings [23], [24]. Trust-calibrated multi-agent decision intelligence suggests that coordination should include mechanisms for assessing the reliability of other agents and adjusting reliance accordingly [17]. In financial risk control, this means that the system should not treat all signals or subordinate policies as equally trustworthy, especially during stress.

6. Governance, Fairness, and Policy Implications

Governance is central to any deployment of hierarchical reinforcement learning in financial risk control. Regulators expect financial institutions to manage model risk, document assumptions, validate performance, and ensure that human oversight can intervene when necessary [15], [16]. A hierarchical system must therefore be accompanied by a governance framework that defines decision rights at each layer. Who approves the upper-layer objectives. Who monitors lower-layer execution. Who can override the system. Who is accountable when the system contributes to a loss. These questions are not technical afterthoughts. They determine whether the system can be used in regulated activities at all.

Fairness is another critical dimension. Financial risk control can affect access to credit, pricing of insurance, treatment of customers, and allocation of capital across communities. Machine learning systems can reproduce or amplify bias through data, model design, and feedback loops [18], [19]. Hierarchical reinforcement learning adds a layer of complexity because fairness constraints may need to be enforced at multiple levels. An upper layer might set fair lending objectives, while a lower layer makes underwriting or pricing decisions. If the lower layer optimizes only for short-term risk reduction, it may undermine fairness goals. If the upper layer imposes rigid constraints, it may reduce the system's ability to distinguish genuine risk differences. Fairness in this context is not a single metric but a governance commitment that requires ongoing monitoring, stakeholder engagement, and contestability.

Policy implications extend beyond individual institutions. If many financial firms adopt similar hierarchical reinforcement learning systems, their interactions could change market dynamics. Herding, correlated strategies, and synchronized deleveraging are possible outcomes. Regulators may need to monitor algorithmic concentration, require circuit breakers, and demand transparency about automated risk controls. Ethical frameworks for artificial

intelligence emphasize transparency, accountability, and human oversight [20], [21]. These principles are compatible with hierarchical reinforcement learning if the architecture is designed to expose decision layers rather than hide them. A policy environment that encourages standardized interfaces, audit trails, and stress-testing protocols could reduce systemic risk while preserving innovation.

7. Deployment, Infrastructure, and Sustainability

Deployment of hierarchical reinforcement learning for financial risk control requires substantial infrastructure. Data pipelines must provide timely, accurate, and traceable information from market feeds, internal systems, and external sources. Feature stores must manage versioning and reproducibility. Simulation environments must be realistic enough to train useful policies but safe enough to avoid catastrophic real-world exploration. Model serving infrastructure must support low-latency inference for lower layers and periodic retraining for upper layers. Human-machine interfaces must allow risk managers to inspect goals, constraints, and exceptions. Without these elements, even a well-designed algorithm will fail in production.

Sustainability is often overlooked in algorithmic risk control, but it matters in several senses. Computational sustainability concerns the energy and cost of continuous training and inference, especially if many agents operate at high frequency [7], [11]. Operational sustainability concerns maintainability: can the system be updated without destabilizing existing controls. Institutional sustainability concerns legitimacy: will regulators, customers, and employees continue to trust the system over time. A hierarchical design can support sustainability by modularizing components so that upper-layer policy changes do not require rebuilding every lower-layer skill. However, modularity also creates integration costs and version-control challenges. If interfaces are poorly specified, the system may accumulate technical debt that becomes a source of risk.

Safety engineering provides useful lessons. Reinforcement learning systems can exhibit reward hacking, unsafe exploration, and unintended side effects [22]. In finance, these risks can translate into financial losses, market disruption, or reputational damage. Deployment should therefore include staged rollout, sandbox testing, kill switches, and human escalation. Lower layers should operate within hard constraints that cannot be overridden by learned policies. Upper layers should be subject to periodic review by risk committees and regulators. The system should log not only actions but also goals, confidence estimates, and constraint satisfaction. These logs are essential for post-mortem analysis and for demonstrating accountability.

8. Evaluation, Validation, and Operational Resilience

Evaluation of hierarchical reinforcement learning in financial risk control cannot rely solely on historical backtests. Backtests are vulnerable to overfitting, survivorship bias, and look-ahead bias. They also assume that the market's response to the agent's actions is unchanged, which is unrealistic for large institutions. Forward testing, paper trading, and controlled live experiments are necessary but expensive and risky. Simulation can help, but simulations must include realistic market impact, liquidity constraints, latency, and adversarial behavior. Evaluation should therefore combine multiple methods: scenario analysis, stress testing, adversarial testing, and ongoing monitoring [10], [14], [16]. The goal is not to prove that the system is always right but to understand the conditions under which it fails.

Operational resilience requires that the system degrade gracefully. If data feeds fail, if a lower-layer agent behaves anomalously, or if an upper-layer goal becomes invalid, the system should fall back to safe defaults. This may mean reverting to rule-based limits, reducing exposure, or pausing automation. Resilience also requires redundancy and separation of duties. A single hierarchical agent should not control all aspects of risk. Multiple independent checks, including human review, should be embedded in the architecture. Research on AI safety emphasizes the importance of corrigibility and oversight [22]. In financial risk control, corrigibility means that the system can be corrected or shut down without causing cascading failures.

Validation should also assess fairness and distributional outcomes. If the system changes pricing or credit allocation, it may have disparate impacts across groups. Fairness-aware evaluation requires disaggregated performance metrics, causal analysis, and stakeholder input [18], [19]. These evaluations should be repeated as markets and populations change. A system that is fair at deployment may become unfair after a regime shift or data drift. Hierarchical reinforcement learning can support fairness by encoding constraints at the upper layer and monitoring lower-layer outcomes, but it cannot guarantee fairness without continuous governance. The evaluation process must therefore be institutionalized, not treated as a one-time certification.

9. Cross-Domain Comparisons and Forward-Looking Perspectives

Hierarchical reinforcement learning has been applied in robotics, games, and navigation, where temporal abstraction and modular skills are natural [1], [2], [6]. Finance differs in important ways. Markets are adversarial, non-stationary, and reflexive. Actions can change the environment, and information is distributed asymmetrically. Unlike a robot in a fixed physical world, a financial agent operates in a system where other agents learn and adapt. This makes hierarchy both more necessary and more difficult. It is more necessary because flat policies struggle with long horizons and conflicting objectives. It is more difficult because the upper layer cannot assume that lower-layer skills remain valid when the market regime changes.

Other domains offer useful analogies. Supply chain risk management uses hierarchical planning to coordinate procurement, inventory, and logistics under disruption. Energy grid control uses layered optimization to balance generation, transmission, and distribution. Healthcare operations use hierarchical scheduling to manage patient flow and resource allocation. In each case, the architecture must balance central coordination with local autonomy, and governance must address safety, fairness, and accountability. Financial risk control can learn from these domains by treating hierarchy as a coordination mechanism rather than just a learning technique [9], [11]. Cross-domain comparisons also highlight the importance of standards, audits, and professional norms in making automated decisions legitimate [20].

Forward-looking perspectives suggest several research directions. One is the development of hierarchical agents that can explain their goals and constraints in natural language for human overseers. Another is the integration of causal inference with hierarchy so that upper layers reason about interventions rather than correlations. A third is the design of multi-agent hierarchies that can coordinate across institutions without centralizing sensitive data. A fourth is the creation of regulatory sandboxes where hierarchical reinforcement learning can be tested under supervision. These directions are ambitious, but they respond to real needs. The future of dynamic financial risk control will likely involve a mix of human judgment,

classical risk models, and learning agents. Hierarchy provides a way to organize that mix without pretending that all decisions belong at the same level.

10. Conclusion

Hierarchical reinforcement learning offers a compelling architectural pattern for dynamic financial risk control because it aligns algorithmic decision-making with the layered structure of financial institutions and regulatory oversight. By separating strategic risk posture from tactical execution, it can manage temporal abstraction, reduce action complexity, and support modular adaptation. However, hierarchy does not eliminate the fundamental challenges of reinforcement learning in financial markets. Regime shifts, adversarial behavior, non-stationarity, and multi-agent interaction can still produce instability, hidden risk, and governance failures. The value of hierarchy depends on whether it is designed as an accountable socio-technical system rather than as a standalone optimizer.

The paper has argued that successful deployment requires robust infrastructure, continuous validation, fairness monitoring, and clear policy frameworks. Upper layers should encode risk appetite, fairness constraints, and regulatory obligations in ways that are transparent and reviewable. Lower layers should operate within hard bounds, with logging and escalation mechanisms that allow human intervention. Multi-agent coordination should be calibrated by trust and reliability rather than assumed. Evaluation should go beyond backtests to include stress scenarios, adversarial testing, and live monitoring. Governance should define decision rights, accountability, and contestability at every level.

Ultimately, hierarchical reinforcement learning for financial risk control is not merely a technical question about value decomposition or option discovery. It is a question about how societies delegate consequential decisions to machines. If designed with care, hierarchy can make automated risk control more adaptable, more auditable, and more resilient. If designed poorly, it can obscure responsibility and amplify systemic risk. The path forward requires interdisciplinary collaboration among computer scientists, financial engineers, regulators, ethicists, and practitioners. The goal should be a layered system that is not only intelligent but also legitimate, sustainable, and safe.

References

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

2. Barto, A. G., & Mahadevan, S. (2003). Recent advances in hierarchical reinforcement learning. Discrete Event Dynamic Systems, 13(4), 341–379.

3. Sutton, R. S., Precup, D., & Singh, S. (1999). Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. Artificial Intelligence, 112(1–2), 181–211.

4. Dietterich, T. G. (2000). Hierarchical reinforcement learning with the MAXQ value function decomposition. Journal of Artificial Intelligence Research, 13, 227–303.

5. Parr, R., & Russell, S. (1998). Reinforcement learning with hierarchies of machines. Advances in Neural Information Processing Systems, 10, 1043–1049.

6. Vezhnevets, A. S., Osindero, S., Schaul, T., Heess, N., Jaderberg, M., Silver, D., & Kavukcuoglu, K. (2017). FeUdal networks for hierarchical reinforcement learning. Proceedings of the 34th International Conference on Machine Learning, 70, 3540–3549.

7. Nachum, O., Gu, S. S., Lee, H., & Levine, S. (2018). Data-efficient hierarchical reinforcement learning. Advances in Neural Information Processing Systems, 31, 3303–3313.

8. Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., & Hassabis, D. (2015). Human-level control through deep reinforcement learning. Nature, 518(7540), 529–533.

9. Heaton, J. B., Polson, N. G., & Witte, J. H. (2017). Deep learning for finance: Deep portfolios. Applied Stochastic Models in Business and Industry, 33(1), 3–12.

10. Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research, 270(2), 654–669.

11. Dixon, M. F., Halperin, I., & Bilokon, P. (2020). Machine learning in finance: From theory to practice. Springer.

12. Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica, 57(2), 357–384.

13. Ang, A., & Bekaert, G. (2002). International asset allocation with regime shifts. The Review of Financial Studies, 15(4), 1137–1187.

14. Jorion, P. (2006). Value at risk: The new benchmark for managing financial risk (3rd ed.). McGraw-Hill.

15. Hull, J. C. (2018). Risk management and financial institutions (5th ed.). Wiley.

16. Basel Committee on Banking Supervision. (2019). Minimum capital requirements for market risk. Bank for International Settlements.

17. Liu, T. (2026). Trust-Calibrated Multi-Agent Decision Intelligence for Financial Risk Management under Market Regime Shifts. Available at SSRN 7307503.

18. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. ACM Computing Surveys, 54(6), Article 115.

19. Barocas, S., Hardt, M., & Narayanan, A. (2019). Fairness and machine learning: Limitations and opportunities. fairmlbook.org.

20. Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds and Machines, 28(4), 689–707.

21. Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society, 3(2), 1–21.

22. Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., & Mané, D. (2016). Concrete problems in AI safety. arXiv:1606.06565.

23. Busoniu, L., Babuska, R., & De Schutter, B. (2008). A comprehensive survey of multiagent reinforcement learning. IEEE Transactions on Systems, Man, and Cybernetics, Part C, 38(2), 156–172.

24. Shoham, Y., & Leyton-Brown, K. (2008). Multiagent systems: Algorithmic, game-theoretic, and logical foundations. Cambridge University Press.