

Hybrid Symbolic-Neural Reasoning Optimization through Reinforcement Learning for Language Models

Emas Nugroho

Department of Computer Science, University of New Hampshire, Durham, NH, USA.
nugroho1981@unh.edu

Rossana Antonini

Department of Computer Science, University of Central Florida, Orlando, FL, USA.
rantonini@ucf.edu

Gaja Hamrol

Department of Electrical Engineering and Computer Science, University of Missouri,
Columbia, MO, USA.
gajahamrol671@missouri.edu

Abstract

Hybrid symbolic-neural reasoning has emerged as a response to the fragility, opacity, and brittle generalization of purely statistical language models. Although transformer-based architectures and large-scale pretraining have enabled impressive linguistic competence and few-shot adaptation, they continue to struggle with reliable multi-step inference, constraint satisfaction, causal consistency, and verifiable action selection. This paper examines hybrid symbolic-neural reasoning optimization through reinforcement learning for language models from a system-level perspective. It argues that the central challenge is not merely improving model accuracy but designing an integrated socio-technical architecture in which neural generation, symbolic verification, reinforcement learning, retrieval, orchestration, and governance operate as mutually constrained subsystems. The analysis considers structural trade-offs among expressiveness, latency, compute cost, interpretability, fairness, robustness, and sustainability. It explores reinforcement learning as an optimization layer that can shape reasoning trajectories, filter plausible but unsound paths, align outputs with human preferences, and coordinate symbolic and neural components under uncertainty. It then discusses infrastructure requirements, deployment patterns, operational sustainability, accountability mechanisms, and policy implications. The paper concludes that hybrid symbolic-neural reinforcement learning systems can improve reasoning performance only if they are treated as governed infrastructures rather than isolated model improvements. Their long-term value depends on auditable reward design, transparent verification pathways, resilient deployment, equitable access, and institutional capacity to manage continuous adaptation.

Keywords

hybrid reasoning, symbolic-neural systems, reinforcement learning, language models, system architecture, AI governance.

1. Introduction

Large language models have become a dominant paradigm in artificial intelligence because they combine broad linguistic coverage with scalable representation learning and flexible generation. The transformer architecture provided a general mechanism for sequence modeling, attention, and parallel training that made very large models feasible [1]. Subsequent work demonstrated that scale, instruction data, and few-shot prompting could produce surprisingly general task behavior without task-specific parameter updates [2]. These advances have shifted the field from narrow supervised pipelines toward foundation models that can be adapted through prompts, retrieval, fine-tuning, and reinforcement learning. However, the same flexibility that makes language models useful also makes them difficult to constrain, verify, and govern. Their reasoning is frequently plausible rather than sound, their confidence poorly calibrated, and their failure modes context-dependent.

Prompting methods such as chain-of-thought reasoning and tree-structured deliberation have improved performance on tasks requiring intermediate steps, arithmetic, commonsense inference, and multi-hop composition [3], [4]. Tool use has further extended language models by allowing them to call external functions, retrieve information, and delegate specialized operations [5]. Yet these techniques often remain fragile because they rely on the model to generate, evaluate, and select its own reasoning paths. A model may produce a coherent explanation for an incorrect conclusion, overuse tools, misread constraints, or fail to maintain consistency across long horizons. In high-stakes domains such as medicine, law, finance, engineering, and public administration, such failures are not merely technical inconveniences. They become accountability problems, safety problems, and legitimacy problems.

Hybrid symbolic-neural reasoning offers a way to combine the pattern recognition and language fluency of neural models with the explicit structure, verifiability, and compositional rigor of symbolic systems. Symbolic components can represent rules, constraints, ontologies, causal relations, and proof obligations. Neural components can parse ambiguous language, generate candidate hypotheses, retrieve relevant context, and translate between informal user intent and formal representations. Reinforcement learning can then serve as an optimization layer that selects among candidate reasoning trajectories, filters paths that violate constraints, and aligns behavior with external feedback. Recent work on scalable reasoning path filtering and alignment for smaller parameter language models illustrates how trajectory-level optimization can improve reasoning efficiency without necessarily relying on enormous model scale [6]. This line of inquiry suggests that future progress may depend less on simply increasing parameter counts and more on designing hybrid architectures that allocate computation intelligently across neural, symbolic, retrieval, and optimization modules.

This paper adopts a system-level perspective. It treats hybrid symbolic-neural reasoning optimization as an infrastructure problem involving architecture, deployment, governance, sustainability, robustness, fairness, and policy. The argument is that reinforcement learning can improve reasoning only when its rewards, environments, verification mechanisms, and operational constraints are designed as part of a larger socio-technical system. The next sections develop this argument by examining conceptual foundations, reinforcement learning as an optimization layer, system architecture, deployment and sustainability, robustness and fairness, and evaluation and policy implications.

2. Conceptual Foundations of Hybrid Symbolic-Neural Reasoning

Symbolic artificial intelligence historically emphasized explicit knowledge representation, logical inference, search, planning, and rule-based manipulation. Its strengths include transparency, compositional structure, and the ability to provide proofs or justifications for

conclusions. Its weaknesses include brittleness in open-world environments, difficulty handling noisy natural language, and high maintenance costs for large knowledge bases. Neural approaches, by contrast, learn distributed representations from data and excel at perception, language, and pattern completion. They can generalize across surface forms and tolerate ambiguity, but they often lack guaranteed consistency, interpretable intermediate states, and reliable mechanisms for enforcing hard constraints. Hybrid systems attempt to exploit the strengths of both traditions while mitigating their respective weaknesses.

Reinforcement learning provides a natural framework for coordinating these components because it focuses on sequential decision-making under uncertainty. In reinforcement learning, an agent learns to select actions that maximize cumulative reward through interaction with an environment [18]. This formulation maps well onto reasoning tasks in which a language model generates intermediate steps, chooses tools, queries symbolic solvers, or decides when to stop. The reasoning process can be treated as a trajectory, and reinforcement learning can optimize the trajectory rather than only the final answer. This is important because many reasoning failures arise from local decisions that appear reasonable but lead to globally invalid conclusions. A model may choose an irrelevant premise, apply an inappropriate rule, or ignore a critical constraint. Trajectory-level optimization can penalize such choices and reward paths that are not only correct but also efficient, auditable, and consistent.

Deep reinforcement learning demonstrated that neural networks can learn complex policies in high-dimensional environments, including games and control tasks. The success of systems such as AlphaGo showed that neural intuition could be combined with tree search and symbolic evaluation to achieve superhuman performance in domains requiring long-term planning [7]. This hybrid pattern is instructive for language model reasoning. A neural policy can propose candidate moves or reasoning steps, while a symbolic or search-based component evaluates their consequences. Reinforcement learning can then improve the proposal policy over time. In language models, the analogous structure involves generating candidate reasoning paths, verifying them with external tools or formal constraints, and updating the policy to favor paths that satisfy multiple criteria.

Human feedback has become a central mechanism for aligning language models with desired behavior. Reinforcement learning from human preferences allows models to learn from comparisons rather than explicit labels, which is useful when quality is difficult to specify programmatically [8]. Instruction tuning with human feedback further demonstrated that reinforcement learning can shape model behavior toward helpfulness, harmlessness, and instruction following [9]. Constitutional and AI-feedback methods explore ways to reduce reliance on direct human labeling by using explicit principles and model-generated critiques [10]. These developments matter for hybrid symbolic-neural reasoning because they show that reinforcement learning can encode normative and procedural preferences, not only task accuracy. However, they also raise governance questions about who defines the reward, how disagreement is handled, and whether optimization pressure creates unintended behavior.

3. Reinforcement Learning as an Optimization Layer for Reasoning

In hybrid symbolic-neural systems, reinforcement learning functions as an optimization layer that mediates among neural generation, symbolic verification, retrieval, and external tools. The neural model can generate a distribution over possible next steps, explanations, or tool calls. The symbolic component can check consistency, enforce constraints, and produce certificates or counterexamples. The reinforcement learning policy can then learn when to trust the neural proposal, when to invoke a solver, when to retrieve additional evidence, and

when to abstain. This architecture treats reasoning as a control problem rather than a single forward pass. The policy must balance exploration of alternative reasoning paths with exploitation of known effective strategies.

A key challenge is reward design. If the reward focuses only on final answer correctness, the policy may learn shortcuts that happen to produce correct outputs for the wrong reasons. Such shortcuts can fail under distribution shift or adversarial conditions. If the reward emphasizes intermediate symbolic validity, the system may become overly conservative and unable to handle ambiguous or incomplete information. If the reward includes efficiency, the model may skip necessary verification. A well-designed reward therefore has to combine multiple objectives, including correctness, consistency, calibration, interpretability, tool-use appropriateness, latency, and cost. Recent work on scalable reasoning path filtering and alignment for smaller parameter language models [6] reflects this multi-objective orientation by emphasizing filtered trajectories and alignment rather than raw generation alone. The lesson is that reinforcement learning is not a neutral optimizer; it is a mechanism for encoding institutional and technical priorities.

Policy optimization methods must also contend with the non-stationary nature of hybrid reasoning environments. The symbolic solver may be updated, retrieval indexes may change, tools may become unavailable, and user populations may shift. A policy that performs well in a controlled benchmark may degrade in deployment. Reinforcement learning can adapt to such changes, but adaptation itself creates risks of reward hacking, drift, and catastrophic forgetting. Governance therefore requires continuous monitoring, versioned policies, staged rollouts, and rollback mechanisms. In this sense, the reinforcement learning layer is less like a one-time training procedure and more like an operational control system that must be maintained over time.

The optimization layer also affects interpretability. Reinforcement learning policies are often opaque, especially when they are neural networks. If a model learns to prefer certain reasoning paths based on subtle reward signals, users may not be able to understand why one path was selected over another. In hybrid systems, interpretability can be improved by requiring symbolic verification artifacts, tool traces, or provenance records. However, such records may reveal only part of the decision process. A central design trade-off is therefore between performance optimization and explanatory accountability. Systems that prioritize auditability may accept lower peak performance, while systems that prioritize aggressive optimization may require stronger external oversight. There is no universal resolution; the appropriate balance depends on domain, risk, and regulatory context.

4. System Architecture and Infrastructure

A hybrid symbolic-neural reasoning system is best understood as a distributed architecture rather than a single model. At its core is a neural language model that handles natural language understanding, generation, and translation into structured representations. Around it are symbolic reasoners, constraint solvers, knowledge graphs, retrieval engines, code interpreters, simulation environments, and verification services. A reinforcement learning controller coordinates these components by selecting actions, allocating compute, and deciding when to stop. An orchestration layer manages state, logging, security, and service-level objectives. This architecture resembles a modern data-intensive application more than a standalone machine learning model.

Infrastructure choices have profound implications. Latency-sensitive applications may require local caching, small distilled models, or edge deployment, while high-stakes analytical tasks may justify remote symbolic solvers and deeper search. Throughput-oriented systems may batch requests, but reasoning tasks often require variable computation per query. Memory management becomes complex because symbolic states, retrieved documents, tool outputs, and intermediate reasoning traces can consume substantial resources. The system must also handle partial failures: a retrieval service may time out, a solver may return an error, or a tool may produce malformed output. Robust orchestration must degrade gracefully rather than collapse into hallucination or silence.

Technical debt is a major concern in such systems. Machine learning systems often accumulate hidden dependencies, feedback loops, data pipeline entanglements, and configuration drift that are difficult to detect until deployment [15]. Hybrid reasoning amplifies this problem because it adds symbolic components, external services, and reinforcement learning policies that interact in complex ways. A change in the reward function may alter tool-use behavior. An update to a knowledge base may invalidate learned policies. A new retrieval index may shift the distribution of prompts. These couplings require strong engineering discipline, version control, reproducible evaluation, and observability. Without such practices, the system may appear to improve in benchmark settings while becoming more fragile in production.

The architecture must also address security and access control. Symbolic solvers and tools may expose sensitive data or provide capabilities that can be misused. Reinforcement learning policies can be manipulated through reward poisoning, environment tampering, or adversarial inputs. Multi-tenant deployments must isolate users, protect proprietary knowledge, and prevent cross-context leakage. Infrastructure therefore includes not only compute and storage but also identity management, encryption, audit logging, and policy enforcement. Hybrid reasoning systems are attractive because they can provide verification, but that verification is only meaningful if the underlying infrastructure is trustworthy.

5. Deployment, Sustainability, and Operational Trade-offs

Deployment of hybrid symbolic-neural reasoning systems involves trade-offs among cost, latency, accuracy, energy, and maintainability. Large language models are computationally expensive to train and serve, and the addition of symbolic solvers, retrieval, and reinforcement learning further increases resource demands. Energy and policy considerations for deep learning in natural language processing have highlighted the environmental costs of large-scale model training and inference [16]. Hybrid systems can sometimes reduce cost by using smaller neural models for routine generation and reserving symbolic or search-based computation for difficult cases. However, such savings depend on accurate routing and efficient orchestration. Poorly designed routing may increase cost by invoking multiple components unnecessarily.

Sustainability also concerns organizational capacity. Continuous reinforcement learning requires data collection, reward modeling, evaluation, and retraining. These processes consume human labor and computational resources. If the system is deployed in a public service, sustainability includes long-term maintenance, staff training, and the ability to respond to changing regulations. If it is deployed in a commercial setting, sustainability includes unit economics and competitive pressure. These factors can encourage premature deployment or excessive optimization for engagement rather than accuracy. A system that is technically capable but institutionally unsupported may fail or cause harm.

Bender and colleagues warned that large language models can produce fluent but misleading text and can encode harmful biases while consuming enormous resources [19]. Hybrid reasoning can mitigate some of these problems by introducing verification and provenance, but it does not eliminate them. Symbolic components may encode outdated or discriminatory rules. Retrieval may surface biased sources. Reinforcement learning may amplify majority preferences. Sustainability therefore has a social dimension: systems must be evaluated not only by whether they work but by whom they serve and at what cost. Deployment decisions should consider accessibility, language coverage, digital divides, and the risk of concentrating computational power in a few institutions.

Operationally, hybrid systems require service-level agreements that reflect reasoning uncertainty. A system may return a verified answer, a tentative answer with caveats, or an abstention. It may request clarification or escalate to a human expert. These behaviors must be designed and governed, not left to the model. Reinforcement learning can be used to calibrate abstention and escalation, but the thresholds should be set through policy deliberation. In high-stakes domains, the cost of a false confident answer may be far greater than the cost of deferring to a human. Deployment patterns should therefore support human-in-the-loop workflows, explainable verification, and reversible actions.

6. Robustness, Fairness, and Governance

Robustness in hybrid symbolic-neural reasoning requires resilience to adversarial inputs, distribution shift, noisy data, and component failures. Adversarial examples may target the neural model, the retrieval system, the symbolic solver, or the reinforcement learning policy. Distribution shift may occur when user behavior changes, new domains emerge, or the knowledge base becomes outdated. Component failures may be accidental or malicious. A robust system must detect anomalies, degrade gracefully, and avoid making unsupported claims. Reinforcement learning can improve robustness by training policies to handle uncertainty and by penalizing overconfident errors, but it can also create new vulnerabilities if the reward is exploited.

Fairness is similarly multifaceted. Neural language models can reproduce social biases present in training data. Symbolic systems can embed discriminatory rules or categories. Reinforcement learning can amplify biases if feedback comes disproportionately from certain groups. Hybrid architectures can improve fairness by enforcing explicit constraints, auditing decision paths, and allowing affected parties to contest outcomes. However, these mechanisms require careful design. A rule that appears neutral may have disparate impact. A verification process may privilege formalizable knowledge over lived experience. A reward model may encode the preferences of a narrow set of annotators. Fairness therefore cannot be reduced to a single metric; it requires ongoing stakeholder engagement and context-sensitive evaluation.

Governance must address the entire lifecycle of the system. This includes problem formulation, data collection, model training, reward design, deployment, monitoring, and decommissioning. Floridi and Cowls proposed a unified framework of principles for AI in society, including beneficence, nonmaleficence, autonomy, justice, and explicability [12]. Mittelstadt and colleagues mapped the ethics of algorithms across concerns such as opacity, bias, privacy, and due process [13]. These frameworks are relevant to hybrid reasoning because reinforcement learning introduces additional opacity and because symbolic components can create an illusion of objectivity. Governance should therefore require

documentation of reward functions, training data, evaluation procedures, known limitations, and appeal mechanisms.

Accountability mechanisms are essential. Raji and colleagues outlined an end-to-end framework for internal algorithmic auditing that connects ethical principles to organizational practice [14]. In hybrid reasoning systems, auditing should examine not only final outputs but also reasoning trajectories, tool invocations, symbolic checks, and reward signals. It should test for bias, robustness, security, and compliance under realistic conditions. Audits should be periodic and triggered by significant changes in policy, data, or infrastructure. External oversight may be necessary for public-sector or high-risk applications. Russell argues that advanced AI systems require careful attention to value alignment and control, especially as capabilities increase [11]. Hybrid symbolic-neural systems do not automatically solve alignment; they create new surfaces for alignment problems because multiple components and objectives must be coordinated.

A further governance challenge is the distribution of power. Hybrid reasoning systems may be expensive to build and operate, which can concentrate capability in large firms, states, or research institutions. Smaller organizations may depend on proprietary APIs, creating lock-in and limiting transparency. Open standards, public compute, and shared evaluation infrastructure can help counter this trend. At the same time, openness can increase misuse risks. Policy must therefore balance access, innovation, security, and accountability. This is not a purely technical decision; it is a political and institutional one.

7. Evaluation, Accountability, and Policy Implications

Evaluation of hybrid symbolic-neural reasoning systems must go beyond accuracy on static benchmarks. It should assess reasoning validity, consistency under perturbation, calibration, robustness, fairness, latency, cost, energy use, and human usability. It should also examine failure modes: when does the system hallucinate, over-rely on tools, ignore symbolic constraints, or fail to abstain? Evaluation should be continuous rather than one-time because reinforcement learning and deployment environments change. It should include red-teaming, adversarial testing, and participatory evaluation with affected communities. In high-stakes settings, evaluation should be tied to regulatory requirements and independent review.

Policy implications are significant. Governments and standards bodies may need to define requirements for transparency, auditability, and human oversight in AI systems that use reinforcement learning and symbolic reasoning. They may also need to address liability when multiple components contribute to harm. If a neural model generates a plausible but false premise, a symbolic solver verifies an invalid formalization, and a reinforcement learning policy selects that path, who is responsible? A governance framework must clarify roles and provide remedies. It must also protect whistleblowers, researchers, and auditors who identify problems. Otherwise, accountability becomes symbolic rather than practical.

Policy should also support sustainable and equitable development. Public investment in compute infrastructure, open datasets, and evaluation benchmarks can reduce concentration and improve reproducibility. Standards for energy reporting and lifecycle assessment can encourage efficiency. Support for interdisciplinary research can bridge computer science, law, ethics, and domain expertise. Regulation should be risk-based and adaptive, avoiding both under-regulation that permits harm and over-regulation that blocks beneficial innovation. The goal is not to slow progress but to ensure that progress is directed toward socially legitimate ends.

Finally, policy must address the global dimension. Language models and hybrid reasoning systems operate across borders, languages, and legal regimes. Their training data and deployment contexts may involve jurisdictions with different norms. International cooperation is needed for safety standards, content governance, and dispute resolution. However, cooperation must respect local values and avoid imposing one region's preferences on others. Hybrid symbolic-neural systems can be designed to accommodate pluralistic rules and value frameworks, but doing so requires deliberate architectural and governance choices. It will not happen automatically through optimization.

8. Conclusion

Hybrid symbolic-neural reasoning optimization through reinforcement learning represents a promising direction for improving the reliability and transparency of language models. By combining neural flexibility with symbolic structure and reinforcement learning control, such systems can generate, verify, and refine reasoning trajectories in ways that pure generation cannot. However, the technical promise is inseparable from system-level challenges. Architecture, infrastructure, deployment, sustainability, robustness, fairness, governance, and policy all shape whether hybrid reasoning systems are safe, equitable, and useful. Reinforcement learning is not a neutral solution; it encodes values through reward design and operational choices. Symbolic components are not automatically objective; they embed assumptions and constraints. Neural components are not automatically general; they reflect data and training conditions.

The central argument of this paper is that hybrid symbolic-neural reasoning should be treated as a governed infrastructure rather than a model feature. Its success depends on auditable reward mechanisms, transparent verification pathways, resilient orchestration, continuous evaluation, and inclusive governance. Smaller, well-aligned models combined with symbolic tools may sometimes outperform larger monolithic models on reasoning tasks, but only if the surrounding system is designed with care. Future research should therefore focus on multi-objective optimization, human-AI escalation, provenance, adversarial robustness, and participatory evaluation. Policy should support these efforts through standards, public infrastructure, and accountability mechanisms. Without such integration, hybrid reasoning may reproduce the same fragility and opacity it seeks to overcome. With it, hybrid symbolic-neural reinforcement learning could contribute to language models that are not only more capable but also more legitimate, contestable, and sustainable.

References

1. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 5998–6008.
2. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877–1901.
3. Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, 35, 24824–24837.

4. Yao, S., Yu, D., Zhao, J., Shafran, I., Griffiths, T. L., Cao, Y., & Narasimhan, K. (2023). Tree of Thoughts: Deliberate problem solving with large language models. *Advances in Neural Information Processing Systems*, 36, 11809–11822.
5. Schick, T., Dwivedi-Yu, J., Dessì, R., Raileanu, R., Lomeli, M., Zettlemoyer, L., Cancedda, N., & Scialom, T. (2023). Toolformer: Language models can teach themselves to use tools. *Advances in Neural Information Processing Systems*, 36, 68539–68551.
6. Li, Q. (2026, May). TrajLite PPO: Scalable Reasoning Path Filtering and Alignment for Small Parameter Large Language Models. In *2026 2nd International Conference on Artificial Intelligence and Digital Ethics (ICAIDE)* (pp. 29-32). IEEE.
7. Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., & Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587), 484–489.
8. Christiano, P. F., Leike, J., Brown, T. B., Martic, M., Legg, S., & Amodei, D. (2017). Deep reinforcement learning from human preferences. *Advances in Neural Information Processing Systems*, 30, 4299–4307.
9. Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Aspell, A., Welinder, P., Christiano, P., Leike, J., & Lowe, R. (2022). Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35, 27730–27744.
10. Bai, Y., Kadavath, S., Kundu, S., Aspell, A., Kernion, J., Jones, A., Chen, A., Goldie, A., Mirhoseini, A., McKinnon, C., Chen, C., Olsson, C., Olah, C., Hernandez, D., Drain, D., Ganguli, D., Li, D., Tran-Johnson, E., Perez, E., ... Kaplan, J. (2022). Constitutional AI: Harmlessness from AI feedback. *arXiv preprint arXiv:2212.08073*.
11. Russell, S. (2019). *Human compatible: Artificial intelligence and the problem of control*. Viking.
12. Floridi, L., & Cows, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review*, 1(1). <https://doi.org/10.1162/99608f92.8cd550d1>
13. Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2). <https://doi.org/10.1177/2053951716679679>
14. Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 33–44). Association for Computing Machinery.
15. Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J.-F., & Dennison, D. (2015). Hidden technical debt in machine learning systems. *Advances in Neural Information Processing Systems*, 28, 2503–2511.

16. Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (pp. 3645–3650). Association for Computational Linguistics.
17. Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., & Mané, D. (2016). Concrete problems in AI safety. arXiv preprint arXiv:1606.06565.
18. Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction (2nd ed.). MIT Press.
19. Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (pp. 610–623). Association for Computing Machinery.