

Reinforcement Learning-Based Evidence Selection for Trustworthy RAG in Financial Analysis and Regulatory Compliance

Beverley J. Miah

School of Electrical Engineering and Computer Science, Oregon State University, Corvallis,
OR, USA.

bmiah@oregonstate.edu

Maximino Almeida

Department of Computer Science, University of Houston, Houston, TX, USA.

maximinoalmeida647@uh.edu

Abstract

Financial analysis and regulatory compliance depend on evidence that is accurate, current, traceable, and defensible under scrutiny. Retrieval-augmented generation has emerged as a practical architecture for grounding language models in external documents, but its trustworthiness in high-stakes financial settings is limited by weak evidence selection, ambiguous provenance, and unstable behavior under distributional shift. This paper examines reinforcement learning-based evidence selection as a system-level mechanism for improving the reliability of retrieval-augmented generation in financial analysis and regulatory compliance. Rather than treating retrieval as a static preprocessing step, the paper frames evidence selection as a sequential decision problem in which an agent learns to choose, sequence, and reject retrieved passages according to task objectives, risk constraints, and governance requirements. The discussion emphasizes architectural trade-offs involving latency, auditability, human oversight, compute sustainability, and fairness. It also considers how reinforcement learning policies can be aligned with regulatory expectations through constrained rewards, uncertainty-aware stopping criteria, and transparent escalation pathways. Case illustrations from earnings analysis, anti-money laundering review, and regulatory reporting show that trustworthy retrieval-augmented generation requires more than improved ranking accuracy. It requires socio-technical infrastructure for provenance, contestability, monitoring, and policy enforcement. The paper concludes that reinforcement learning-based evidence selection can strengthen financial artificial intelligence only when embedded within governance frameworks that treat evidence quality, uncertainty, and accountability as first-class design concerns.

Keywords

retrieval-augmented generation; reinforcement learning; evidence selection; financial analysis; regulatory compliance; trustworthy artificial intelligence; governance; robustness; fairness; socio-technical systems.

1. Introduction

Financial analysis and regulatory compliance are evidence-intensive activities. Analysts, auditors, risk officers, and compliance teams must interpret filings, earnings call transcripts, internal policies, transaction records, supervisory guidance, and market disclosures. The

increasing use of large language models in these domains has created both opportunity and risk. These models can summarize, compare, and explain complex material, but they can also generate plausible yet unsupported claims. Retrieval-augmented generation addresses this problem by conditioning generation on retrieved external text, thereby reducing reliance on parametric memory alone [1]. In practice, however, retrieval-augmented generation does not automatically produce trustworthy outputs. The system may retrieve outdated passages, overlook contradictory evidence, privilege superficially similar documents, or fail to disclose uncertainty when evidence is weak.

Trustworthiness in financial artificial intelligence is therefore not merely a model accuracy problem. It is a socio-technical problem involving document provenance, access control, institutional accountability, regulatory expectations, and the ability to reconstruct decision pathways. A compliance officer needs to know not only what the model answered but also which evidence supported the answer, which evidence was excluded, how conflicts were resolved, and whether the system behaved consistently across comparable cases. Conventional retrieval pipelines optimize relevance, but relevance is not equivalent to evidentiary sufficiency. A passage can be semantically similar to a query while being legally outdated, jurisdictionally inapplicable, or factually superseded. This gap motivates reinforcement learning-based evidence selection, in which an agent learns policies for selecting, sequencing, and rejecting evidence according to task-specific and risk-sensitive objectives.

Reinforcement learning offers a framework for sequential decision-making under uncertainty [9]. In the context of retrieval-augmented generation, evidence selection can be modeled as a trajectory of retrieval actions, sufficiency assessments, and stopping decisions. The agent may retrieve additional passages, reformulate queries, compare sources, or decide that the available evidence is insufficient. Such a policy can be trained using task outcomes, human feedback, regulatory constraints, or simulated review processes. However, deploying reinforcement learning in financial analysis introduces new governance questions. Rewards may encode institutional preferences that are difficult to justify, policies may be opaque to auditors, and optimization may create unintended incentives to suppress uncertainty or overstate confidence. This paper argues that reinforcement learning-based evidence selection can improve trustworthy retrieval-augmented generation only when it is designed as part of a broader accountable infrastructure.

2. Retrieval-Augmented Generation and the Evidence Selection Problem

Retrieval-augmented generation combines a retriever with a generator so that responses are grounded in external documents [1]. Early systems demonstrated that dense passage retrieval could substantially improve open-domain question answering by mapping queries and passages into a shared semantic space [2]. Subsequent work integrated retrieval into language model pre-training and generation, showing that external memory can improve factual coverage and domain adaptation [3]. Generative approaches further refined the interaction between retrieval and answer synthesis, allowing models to condition on multiple passages while generating responses [4]. Large-scale retrieval systems have also demonstrated that language models can benefit from access to trillions of tokens, although the cost and complexity of maintaining such indexes are substantial [5].

Despite these advances, evidence selection remains a persistent bottleneck. Retrieval systems typically return a ranked list of passages, but generation quality depends on which passages are actually used and how conflicts among them are handled. Self-reflective and adaptive

retrieval methods have attempted to improve this process by teaching models to decide when retrieval is necessary and how to critique retrieved evidence [6]. Active retrieval approaches similarly allow models to issue follow-up queries when initial evidence is insufficient [7]. Surveys of retrieval-augmented generation describe a rapidly expanding design space that includes query rewriting, multi-hop retrieval, reranking, compression, and verification [8]. These developments show that retrieval is not a single action but a policy problem involving multiple decisions over time.

In financial analysis, the evidence selection problem is especially pronounced. A query about revenue recognition may require the system to distinguish between generally accepted accounting principles, company-specific accounting policies, auditor notes, and management commentary. A compliance query about suspicious transactions may require comparing internal escalation rules with jurisdictional reporting obligations. In such settings, the most semantically similar passage may not be the most authoritative. The system must weigh source hierarchy, temporal validity, jurisdictional scope, and institutional policy. A retrieved passage from an older regulatory bulletin may be misleading if a newer rule has superseded it. A passage from a news article may be useful for context but inappropriate as primary evidence for a compliance determination.

Evidence selection also involves rejection. Trustworthy systems must be able to say that the available evidence is insufficient, contradictory, or outside the system’s competence. This capability is difficult to achieve with static ranking alone. Reinforcement learning can model rejection as a legitimate action with its own reward structure, allowing the system to avoid forced answers when retrieval confidence is low. The challenge is to design rewards that penalize unsupported confidence without creating excessive conservatism that renders the system unusable. This tension between coverage and caution is central to trustworthy retrieval-augmented generation in financial environments.

3. Reinforcement Learning for Evidence Selection

Reinforcement learning provides a formal framework for learning sequential decisions through interaction with an environment [9]. In deep reinforcement learning, neural networks approximate policies or value functions, enabling agents to act in complex, high-dimensional spaces [10]. Policy gradient methods such as proximal policy optimization have become widely used because they can stabilize training and support continuous improvement [11]. Preference-based reinforcement learning further allows agents to learn from human comparisons rather than hand-coded reward functions [12]. These techniques underpin modern alignment methods for language models, including instruction tuning with human feedback and preference optimization [13]. Earlier work on fine-tuning language models from human preferences established that human judgments can shape generation behavior in nuanced ways [14]. Safety research has also highlighted that optimizing proxy rewards can produce unintended behavior, including reward hacking and negative side effects [15].

For evidence selection, reinforcement learning can be applied at several levels. At the retrieval level, an agent can decide whether to issue an initial query, reformulate it, or expand the search space. At the passage level, the agent can select a subset of retrieved documents, order them for generation, or discard low-authority sources. At the generation level, the agent can decide whether to answer, request clarification, or escalate to a human reviewer. These decisions form a trajectory whose reward may include answer correctness, citation precision, regulatory compliance, latency, and cost. The design of this reward is a governance act, not merely a technical parameter. If the reward prioritizes rapid answers, the system may learn to

avoid additional retrieval. If the reward penalizes unsupported claims strongly, the system may become overly cautious and defer too many cases to human review.

Financial applications require risk-sensitive rewards. A compliance system may assign asymmetric costs to false negatives and false positives. Missing a suspicious transaction may be far more costly than flagging a legitimate one for review. In earnings analysis, misstating a material risk may have reputational and legal consequences. Reinforcement learning can incorporate such asymmetries, but this requires explicit institutional deliberation about acceptable error profiles. The policy may also need to satisfy hard constraints, such as never using privileged information outside authorized access controls or never citing documents from an excluded jurisdiction. Constrained reinforcement learning and shielding mechanisms can enforce such boundaries, but they add architectural complexity and may reduce policy flexibility.

Another challenge is credit assignment. In a multi-step evidence selection process, it is difficult to determine whether a poor final answer resulted from bad retrieval, poor ranking, inappropriate synthesis, or an unrealistic query. Reward shaping and process supervision can help, but they introduce additional assumptions. Human feedback can provide richer signals, but human reviewers may disagree, and their judgments may reflect institutional biases. For trustworthy financial systems, reinforcement learning must therefore be paired with detailed logging, counterfactual analysis, and periodic policy audits. The learned policy should be treated as an auditable institutional artifact rather than an opaque optimizer.

4. System Architecture and Deployment in Financial Analysis

A practical architecture for reinforcement learning-based evidence selection includes ingestion, indexing, retrieval, selection, generation, verification, and audit layers. The ingestion layer must handle heterogeneous documents, including regulatory filings, internal policies, transaction records, and market data. It must preserve metadata such as source, date, jurisdiction, access level, and version history. The indexing layer must support efficient semantic and keyword retrieval while maintaining permission-aware search. The retrieval layer produces candidate evidence, and the selection layer applies a learned policy to choose, order, or reject passages. The generation layer synthesizes an answer with citations, while the verification layer checks claims against selected evidence. The audit layer records the full trajectory, including queries, candidate sets, selected passages, policy decisions, and human interventions.

Uncertainty-aware retrieval-augmented generation is particularly relevant to this architecture. Recent work on uncertainty-aware retrieval-augmented generation emphasizes trustworthy generation and risk control, suggesting that systems should estimate when evidence is insufficient and adjust behavior accordingly [16]. In financial analysis, uncertainty may arise from sparse retrieval results, conflicting sources, ambiguous language, or temporal gaps. A reinforcement learning policy can learn to respond to uncertainty by retrieving additional evidence, requesting human review, or providing a qualified answer with explicit limitations. The architecture must support these behaviors without creating excessive latency. A compliance officer may tolerate a slower response if the system provides a defensible evidence trail, while a trading analyst may require faster answers with clear confidence indicators.

Deployment considerations extend beyond model performance. Financial institutions operate under strict access controls, data residency requirements, and records retention policies. A

retrieval-augmented generation system must respect these constraints at every layer. The reinforcement learning policy cannot simply optimize for answer quality if doing so violates permission boundaries or leaks sensitive information across business units. Deployment also requires integration with existing case management, surveillance, and reporting systems. If the system cannot export evidence trails into these systems, its outputs may be difficult to use in formal compliance processes. Human-in-the-loop workflows are therefore essential. The system should know when to escalate, how to present conflicting evidence, and how to capture reviewer feedback for future policy improvement.

Operational sustainability is another architectural concern. Maintaining large indexes, running frequent retrieval, and fine-tuning reinforcement learning policies consume significant compute resources. In a financial institution, these costs must be weighed against risk reduction and analyst productivity. A policy that retrieves many passages to maximize accuracy may be unsustainable at scale. A policy that retrieves too few may be unreliable. The architecture should therefore support adaptive retrieval budgets, caching, and tiered review. Trustworthy systems need not maximize evidence collection in every case; they need to allocate evidentiary effort according to risk, materiality, and user role. This allocation is itself a policy decision that should be monitored and governed.

5. Trustworthy RAG, Governance, and Regulatory Compliance

Trustworthy retrieval-augmented generation in finance cannot be separated from governance. Financial institutions are accustomed to model risk management, audit trails, and regulatory examinations. An evidence selection policy should be documented, versioned, and validated like other decision systems. Governance bodies should define acceptable evidence sources, source hierarchies, conflict resolution rules, and escalation thresholds. They should also specify how the system handles uncertainty. If the policy learns to suppress uncertainty because doing so improves a reward metric, the institution may face compliance failures. Governance must therefore review not only outcomes but also the reward design, training data, and policy behavior across demographic, geographic, and product segments.

Regulatory technology, often called RegTech, has emerged as a domain where automation can improve compliance efficiency while preserving stability [17]. However, financial innovation can also create regulatory arbitrage and new forms of opacity [18]. Reinforcement learning-based evidence selection may improve consistency by applying documented policies across large volumes of cases, but it may also obscure the reasoning behind decisions if the policy is not explainable. Regulators increasingly expect financial institutions to explain automated decisions, retain records, and demonstrate effective challenge. A retrieval-augmented generation system should therefore produce not only answers but also evidence maps, confidence indicators, and audit logs. These artifacts allow reviewers to reconstruct why a particular conclusion was reached and whether the selected evidence was appropriate.

Fairness and bias are also central to trustworthy governance. Machine learning systems can perpetuate historical biases present in training data or institutional practices [19]. In financial compliance, bias may manifest in uneven scrutiny of certain customer groups, languages, or transaction patterns. An evidence selection policy might systematically overlook documents written in minority languages or prioritize sources that reflect dominant institutional perspectives. Fairness cannot be reduced to a single metric. It requires ongoing monitoring, stakeholder engagement, and explicit trade-offs between competing objectives [20]. Governance should require periodic fairness audits and mechanisms for affected parties to

contest decisions. In regulatory compliance, contestability is not merely an ethical ideal; it is often a legal and procedural requirement.

The governance of reinforcement learning-based evidence selection must also address accountability. If the policy makes a decision that leads to a compliance breach, who is responsible? The model developer, the data provider, the compliance officer, or the institution? Clear accountability chains are necessary. Human oversight should not be a symbolic checkbox. Reviewers need enough information to challenge the system and enough authority to override it. The system should record overrides and use them as learning signals, but it should also protect against feedback loops that simply reinforce existing institutional preferences. A trustworthy system is one that can be questioned, corrected, and improved without undermining its operational value.

6. Robustness, Fairness, Sustainability, and Infrastructure Trade-Offs

Robustness is a major requirement in financial applications. Markets change, regulations evolve, and adversaries may attempt to manipulate documents or queries. A reinforcement learning-based evidence selection policy trained on historical data may fail when new regulations are introduced or when market conditions shift. Robustness requires continuous monitoring, out-of-distribution detection, and rapid retraining or policy adjustment. It also requires adversarial testing. A malicious actor might inject misleading documents into a retrieval corpus or craft queries designed to elicit unsupported claims. The system should be able to detect anomalies in retrieved evidence and escalate when source integrity is uncertain. Robustness is not a static property; it is an ongoing operational discipline.

Fairness intersects with robustness in complex ways. A policy that performs well on average may perform poorly for specific jurisdictions, languages, or customer segments. For example, a system trained primarily on English-language filings may mishandle multilingual compliance documents. A policy optimized for large corporate disclosures may not generalize to small business lending. Fairness audits should therefore examine performance across relevant groups and contexts. They should also consider allocative harms, such as disproportionate scrutiny, and representational harms, such as stereotyping in generated summaries. Addressing these issues requires diverse training data, inclusive evaluation, and governance structures that include affected communities.

Sustainability is often overlooked in discussions of trustworthy artificial intelligence. Large-scale retrieval and reinforcement learning require significant energy and hardware resources. The environmental cost of training and deploying these systems must be weighed against their benefits. In financial institutions, sustainability may also mean economic sustainability. A system that requires constant expert review may not scale. A system that creates excessive false positives may burden compliance teams. A system that is too slow may be abandoned by analysts. Sustainable design involves creating feedback loops that improve performance over time while controlling operational costs. It also involves selecting evidence sources that are stable, authoritative, and maintainable rather than relying on brittle web-scale retrieval.

Infrastructure trade-offs are unavoidable. Centralized indexes can improve consistency but create single points of failure and raise data governance concerns. Federated retrieval can preserve jurisdictional control but complicates evidence comparison. Real-time retrieval can improve freshness but increases latency and cost. Batch retrieval can support auditability but may miss time-sensitive developments. Reinforcement learning policies can adapt to these trade-offs, but they cannot eliminate them. The institution must decide which trade-offs are

acceptable for which use cases. A trading desk may prioritize latency, while a compliance review may prioritize traceability. A mature architecture supports multiple operating modes with clear governance rules for each.

7. Evaluation, Case Illustrations, and Cross-Domain Comparisons

Evaluation of reinforcement learning-based evidence selection must go beyond answer accuracy. Financial applications require measures of citation correctness, evidence sufficiency, source authority, temporal validity, and conflict resolution. A system that produces the right answer for the wrong reason may be untrustworthy because it cannot be defended in an audit. Evaluation should therefore include process metrics, such as whether the policy retrieved authoritative sources before generating a conclusion. It should also include robustness tests, such as performance under contradictory evidence or missing documents. Human evaluation remains important, but it must be structured to capture inter-reviewer disagreement and institutional standards. Automated evaluation can supplement human review, but it should not replace accountability.

Consider an earnings analysis scenario. An analyst asks why a company's operating margin declined. A retrieval-augmented generation system might retrieve the latest earnings release, management commentary, and analyst reports. A reinforcement learning policy could learn to prioritize the earnings release and audited financial statements over secondary commentary. It might also detect that the analyst report relies on an older forecast and therefore should not be used as primary evidence. If evidence is insufficient, the policy could retrieve the previous quarter's filing for comparison. The final answer should include citations and a note about uncertainty. The system's value lies not only in the answer but in the evidence trail that allows the analyst to verify the reasoning.

In anti-money laundering review, the stakes are higher. A compliance analyst may ask whether a set of transactions meets the threshold for a suspicious activity report. The system must retrieve internal policies, jurisdictional regulations, and case precedents. A reinforcement learning policy might learn to weigh regulatory text more heavily than internal guidance, while still considering institutional risk appetite. If regulations conflict across jurisdictions, the policy should escalate rather than choose arbitrarily. The system must also avoid disparate impact. If the policy systematically flags transactions from certain regions without sufficient evidence, it may create fairness and legal problems. Evaluation must therefore include fairness metrics alongside accuracy and efficiency.

Cross-domain comparisons are instructive. In healthcare, retrieval-augmented generation must handle clinical guidelines, patient records, and privacy regulations. In legal research, it must distinguish binding precedent from persuasive authority. In financial analysis, it must handle accounting standards, market disclosures, and supervisory guidance. Each domain has its own source hierarchy and temporal dynamics. Reinforcement learning policies can be adapted across domains, but the reward functions and governance structures cannot be copied uncritically. A policy that works in open-domain question answering [1] may fail in regulatory compliance. A self-reflective retrieval method [6] may improve general robustness but still require domain-specific constraints. Active retrieval strategies [7] may reduce missing evidence but increase latency. Uncertainty-aware approaches [16] provide a useful foundation, but they must be embedded in domain-specific accountability systems.

8. Policy Implications and Forward-Looking Perspectives

Policymakers and regulators face a difficult task. They must encourage innovation in financial artificial intelligence while protecting market integrity, consumer interests, and systemic stability. Reinforcement learning-based evidence selection could improve compliance if it makes decisions more consistent and evidence-based. However, it could also create new forms of opacity if institutions cannot explain how policies were trained or why particular evidence was selected. Regulatory frameworks should therefore emphasize process transparency, documentation, and contestability rather than prescribing specific algorithms. Institutions should be required to demonstrate that their systems can identify insufficient evidence, escalate appropriately, and produce auditable records. They should also be required to test for bias and robustness across relevant scenarios.

Industry standards could play a constructive role. Shared ontologies for evidence provenance, common formats for audit trails, and certification regimes for trustworthy retrieval systems could reduce fragmentation. RegTech collaborations [17] could help smaller institutions adopt safe practices without building every component from scratch. However, standards must avoid locking in proprietary approaches or creating barriers to entry. They should also address regulatory arbitrage, where firms might deploy opaque systems in less supervised jurisdictions [18]. International coordination is necessary because financial markets and data flows cross borders. A system that is trustworthy in one jurisdiction may be non-compliant in another if it mishandles data residency or disclosure rules.

Research priorities should include methods for interpretable reinforcement learning policies, uncertainty quantification in evidence selection, and fairness-aware reward design. Interpretability is essential for governance. A policy that cannot be explained may be difficult to validate or contest. Uncertainty quantification can help the system recognize when it lacks sufficient evidence and should defer to humans. Fairness-aware reward design can mitigate disparate impacts, but it requires careful specification of protected attributes and contextual harms [19]. Researchers should also study how human feedback shapes institutional behavior. If reviewers consistently favor certain sources or outcomes, the policy may amplify those preferences [20]. Governance mechanisms must detect and correct such drift.

Forward-looking perspectives suggest that trustworthy retrieval-augmented generation will become a core infrastructure for financial analysis and compliance. The most successful systems will not be those that simply achieve high benchmark scores. They will be those that integrate retrieval, reinforcement learning, uncertainty management, and governance into a coherent socio-technical architecture. They will support human expertise rather than replace it. They will make evidence selection visible, contestable, and accountable. They will balance accuracy with fairness, robustness, and sustainability. The challenge is substantial, but the potential is equally significant. Financial institutions that invest in trustworthy evidence infrastructure may gain not only efficiency but also resilience in an increasingly complex regulatory environment.

9. Conclusion

Reinforcement learning-based evidence selection offers a promising approach to improving trustworthy retrieval-augmented generation in financial analysis and regulatory compliance. By treating evidence selection as a sequential decision problem, systems can learn when to retrieve, which sources to prioritize, when to reject insufficient evidence, and when to escalate to human review. This capability addresses a central weakness of conventional retrieval-augmented generation, which often optimizes relevance without considering authority, temporal validity, conflict, or uncertainty. However, reinforcement learning does not

automatically produce trustworthy outcomes. Reward design, policy transparency, governance, fairness, and sustainability are essential components of any responsible deployment.

The paper has argued that trustworthy retrieval-augmented generation must be understood as a socio-technical system. Architecture, infrastructure, regulation, and institutional accountability are not external constraints but constitutive elements of trust. Financial institutions should therefore invest in audit trails, uncertainty-aware policies, human oversight, and fairness monitoring. Regulators should encourage process transparency and contestability without stifling innovation. Researchers should develop methods that make learned evidence selection explainable and governable. Ultimately, the value of reinforcement learning in this domain will depend less on algorithmic novelty and more on whether it can be embedded within institutions that treat evidence, uncertainty, and accountability as first-class concerns.

References

1. Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. In *Advances in Neural Information Processing Systems* (Vol. 33, pp. 9459-9474). Curran Associates.
2. Karpukhin, V., Oğuz, B., Min, S., Lewis, P., Wu, L., Edunov, S., Chen, D., & Yih, W. (2020). Dense passage retrieval for open-domain question answering. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing* (pp. 6769-6781). Association for Computational Linguistics.
3. Guu, K., Lee, K., Tung, Z., Pasupat, P., & Chang, M. (2020). Retrieval augmented language model pre-training. In *Proceedings of the 37th International Conference on Machine Learning* (Vol. 119, pp. 3929-3938). PMLR.
4. Izacard, G., & Grave, E. (2021). Leveraging passage retrieval with generative models for open domain question answering. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 874-880). Association for Computational Linguistics.
5. Borgeaud, S., Mensch, A., Hoffmann, J., Cai, T., Rutherford, E., Millican, K., van den Driessche, G., Lespiau, J.-B., Damoc, B., Clark, A., de Las Casas, D., Guy, A., Menick, J., Ring, R., Hennigan, T., Huang, S., Maggiore, L., Jones, C., Cassirer, A., ... Sifre, L. (2022). Improving language models by retrieving from trillions of tokens. In *Proceedings of the 39th International Conference on Machine Learning* (Vol. 162, pp. 2206-2240). PMLR.
6. Asai, A., Wu, Z., Wang, Y., Sil, A., & Hajishirzi, H. (2024). Self-RAG: Learning to retrieve, generate, and critique through self-reflection. In *The Twelfth International Conference on Learning Representations*. OpenReview.net.
7. Jiang, Z., Xu, F. F., Gao, L., Sun, Z., Liu, Q., Dwivedi-Yu, J., Yang, Y., Callan, J., & Neubig, G. (2023). Active retrieval augmented generation. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing* (pp. 7969-7992). Association for Computational Linguistics.

8. Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., & Wang, H. (2023). Retrieval-augmented generation for large language models: A survey. arXiv. <https://arxiv.org/abs/2312.10997>
9. Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction (2nd ed.). MIT Press.
10. 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.
11. Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. arXiv. <https://arxiv.org/abs/1707.06347>
12. Christiano, P. F., Leike, J., Brown, T., Martic, M., Legg, S., & Amodei, D. (2017). Deep reinforcement learning from human preferences. In *Advances in Neural Information Processing Systems* (Vol. 30, pp. 4299-4307). Curran Associates.
13. 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., Askell, A., Welinder, P., Christiano, P., Leike, J., & Lowe, R. (2022). Training language models to follow instructions with human feedback. In *Advances in Neural Information Processing Systems* (Vol. 35, pp. 27730-27744). Curran Associates.
14. Ziegler, D. M., Stiennon, N., Wu, J., Brown, T. B., Radford, A., Amodei, D., Christiano, P., & Irving, G. (2019). Fine-tuning language models from human preferences. arXiv. <https://arxiv.org/abs/1909.08593>
15. Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., & Mané, D. (2016). Concrete problems in AI safety. arXiv. <https://arxiv.org/abs/1606.06565>
16. Zhang, R. (2025, December). Uncertainty-Aware RAG: Trustworthy Generation and Risk Control. In *2025 5th International Symposium on Artificial Intelligence and Big Data (AIBDF)* (pp. 526-531). IEEE.
17. Arner, D. W., Barberis, J. N., & Buckley, R. P. (2017). FinTech and RegTech: Enabling innovation while preserving financial stability. *Georgetown Journal of International Affairs*, 18(3), 47-58.
18. Buchak, G., Matvos, G., Piskorski, T., & Seru, A. (2018). Fintech, regulatory arbitrage, and the rise of shadow banks. *Journal of Financial Economics*, 130(3), 453-483.
19. 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.
20. Barocas, S., Hardt, M., & Narayanan, A. (2019). Fairness and machine learning: Limitations and opportunities. fairmlbook.org.